

2019년 봄호

연구방법논총

Journal of Research Methodology

【연구논문】

- 사회과학 초보 연구자의 질적 연구 수행경험 : 사전연구 경험을 바탕으로
유전향 · 김유진 · 서미경 / 1
- 휴대폰 사용은 조현병 환자에게 지각된 사회지지를 반영하는 지표가 될 수 있나? 우리남 · 최성진 / 21
- 공공부문 프로그램평가를 위한 평가역량구축 방안의 모색 제갈돈 / 35
- 장애인 거주시설 종사자의 인권기반 복지실천 경험에 관한 연구 전지혜 / 61
- 공공기관 고객만족지수(PCS)모형을 활용한 기록정보서비스 품질조사: P기록관을 중심으로
박제우 · 윤유라 / 93

【연구경향】

- An Introduction to Statistical Causal Inference : Concepts, Models and Some Methodological Issues
Byung-jae Lee / 111
- Indeterminate Feature of Parameter Estimation in Multilevel Categorical Latent Variable Jungkyu Park / 151



경북대학교 사회과학기초자료연구소

차 례

【연구논문】

- 사회과학 초보 연구자의 질적 연구 수행경험: 사전연구 경험을 바탕으로
..... 유전양·김유진·서미경 / 1
- 휴대폰 사용은 조현병 환자에게 지각된 사회지지를 반영하는 지표가 될 수 있나?
..... 우리님·최성진 / 21
- 공공부문 프로그램평가를 위한 평가역량구축 방안의 모색 제갈돈 / 35
- 장애인 거주시설 종사자의 인권기반 복지실천 경험에 관한 연구 전지혜 / 61
- 공공기관 고객만족지수(PCSI)모형을 활용한 기록정보서비스 품질조사: P기록관을
중심으로 박제우·윤유라 / 93

【연구경향】

- An Introduction to Statistical Causal Inference: Concepts, Models and
Some Methodological Issues Byung-Jae Lee / 111
- Indeterminate Feature of Parameter Estimation in Multilevel Categorical
Latent Variable Jungkyu Park / 151
- 『연구방법논총』 원고작성의 일반적 요령 175
- 『연구방법논총』 저술 윤리강령 182
- 『연구방법논총』 편집 및 심사 규정 184
- 『연구방법논총』 편집위원명단 188

Journal of Research Methodology

| 2019 Spring Vol.4, No.1 |

- Experience of conducting qualitative research by a novice researcher in social science : Based on practicing the pilot study
Chuangyang Yu · Yujin Kim · Mikyung Seo / 1
- Can Use of Mobile Phone Be Applied as an Indicator of Perceived Social Supports in Patients with Schizophrenia?
Rinim Woo · SeongJin Choi / 21
- Evaluation Capacity Building for Program Evaluation in Public Sector
Don Jaegal / 35
- A Qualitative Case Study on the Human Rights based Approach in Living Facility of Persons with Disabilities: Focusing on the Experiences of Social Workers
Jihye Jeon / 61
- Measuring of Recording Information Service Quality Using the PCSI(Public Customer Satisfaction Index) :Case of the P Archives
Je-woo Park · You-Ra Youn / 93
- An Introduction to Statistical Causal Inference: Concepts, Models and Some Methodological Issues
Byung-Jae Lee / 111
- Indeterminate Feature of Parameter Estimation in Multilevel Categorical Latent Variable
Jungkyu Park / 151



KYUNGPPOOK NATIONAL UNIVERSITY
The Institute for Basic Social Science

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.1>

【연구논문】

사회과학 초보 연구자의 질적 연구 수행경험:
사전연구 경험을 바탕으로

유전양*·김유진**·서미경***

논문요약

본 연구의 목적은 연구방법론 차원에서 사전연구에 대해 소개하고, 사전연구를 수행하면서 깨닫게 된 질적 연구 진행 과정상의 어려움과 방법론적인 문제들을 구체적으로 살펴보는 것이다. 사전연구란 주 연구(main study)를 준비하기 위해 실시하는 소규모의 방법론적 실험이며, 연구자가 관심을 갖고 있는 주제가 현실적으로 수행 가능한 연구 질문이 될 수 있는지 파악하기 위한 예비적 연구이다. 사전연구는 그 유용성에도 불구하고, 질적 연구 방법론 교재에서 충분히 논의되지 않고 있다. 본 연구를 통해 중국 유학생인 초보 연구자의 실제 경험을 바탕으로 질적 연구를 수행하는 데 있어 사전연구의 유용성 및 가치를 강조하고자 하였다.

주제어: 사전연구, 질적 연구방법, 연구 도구로서의 연구자, 연구 방법론

I. 서론

사회과학 분야에서 질적 연구의 활용이 점점 많아지고 있다(김인숙, 2016; Creswell, 2013). 질적 연구는 연구자가 자료수집과 분석의 주요 도구로서 연구 참여

* 경북대학교 대학원 사회복지학부 박사과정생.

** 경북대학교 사회복지학부 부교수, 교신저자.

*** 경북대학교 대학원 사회복지학부 박사수료생.

자의 내부자적 관점을 통해 어떤 현상 및 경험을 심층적으로 탐색하고 이해하기 위한 목적으로 활용하는 연구방법이다(김인숙, 2016; 윤택림, 2013; Creswell, 2013). 질적 연구방법에 대한 높은 관심에도 불구하고, 실제로 연구를 어떻게 진행하는지에 대해서는 막연히 여기는 경우가 많다(이현서·박선웅, 2018). 질적 연구방법을 통해 얻은 결과물이 얼마나 진정성 있고 과학적이며 가치 있는 결과물인지는 연구 과정 그 자체에 달려있다는 점을 고려할 때(Creswell, 2015; Padgett, 1998; Yin, 2011), 연구 진행 과정을 막막하게 여기는 것은 심각한 문제이다.

본 연구는 초보 연구자의 질적 연구 수행 경험을 다루고 있다. 보다 구체적으로 초보 연구자가 연구(main study)를 시작하기 전에 수행한 사전연구(pilot study)의 경험을 바탕으로, 질적 연구 진행 과정상의 어려움과 방법론적인 문제들을 나누고자 한다. 본 연구의 주저자는 외국인 유학생으로서 다문화 사회에 대해 연구하며, 특히 한국사회에 정착한 결혼이민여성의 대학교육과 취업에 관해 특별한 관심을 갖고 있다. 이 주제에 관심을 갖는 이유는 결혼이민여성들이 한국 사회의 소수자이자 사회복지 정책의 수혜자에 머무는 것이 아니라, 자기 삶의 주체이자 당당한 시민으로서 한국사회에 자리매김하도록 어떻게 도울 것인가에 대해 고민하기 때문이다. 연구자는 이와 같은 문제의식을 갖고 관련 주제를 학술적으로 연구하는 방법으로서 질적 연구방법에 관심을 갖게 되었다.

사전연구란 주 연구(main study)를 준비하기 위해 실시하는 소규모의 방법론적 실험이며, 연구자가 관심을 갖고 있는 주제가 현실적으로 수행 가능한 연구 질문이 될 수 있는지 파악하기 위한 예비적 연구라고 할 수 있다(Jariath et al., 2000; Kim, 2011; Prescott and Soeken, 1989; van Teijlingen and Hundley, 2002). 사전연구는 그 유용성에도 불구하고, 질적 연구 방법론 교재에서 충분히 논의되지 않고 있다(Morrison, Clement, Nestel and Brown, 2016). 질적 연구는 실제 연구가 진행되면서 여러 가지 것들이 조율되는 특성 상, 연구자가 융통성을 발휘할 수 있는 유연함을 갖추고 있으면 사전연구를 굳이 하지 않아도 된다고 보는 입장도 있다(Holloway, 2008). 간호학자로서 다수의 질적 연구 방법론 교재를 집필한 Morse(1997)도 질적 연구를 하는 데 사전연구가 굳이 필요하지 않고 사전연구가 오히려 실제 연구에 부정적인 영향을 끼칠 수 있다고 주장하였다. 세부적인 연구방법론(예, 현상학 연구, 근거

이론 등에 따라 다르지만, 일반적으로 질적 연구에서는 자료가 포화상태(saturation)에 이를 때까지 불확실성을 내포하고 있는데, 사전연구를 하면서 연구자가 깊은 탐색도 하기 전에 잠정적 결론을 내리는 오류를 범할 수 있다고 보기 때문이다.

사전연구가 불필요하거나 오히려 방해가 될 수 있다는 이와 같은 의견에도 불구하고, 많은 연구자들은 사전연구의 필요성을 주장한다. 특히 초보 연구자는 필히 사전연구를 수행해야 하는데, 본격적인 주 연구를 시작하기 전에 자료 수집과 분석 방법을 연습하고 평가할 필요가 있기 때문이다(van Teijlingen and Hundley, 2002; Kim, 2011; Leon et al, 2011). 또한, 사전연구를 하면서 실제 본 연구(main study)가 진행될 때 발생할 수 있는 문제들을 예측하여 본 연구에 변경 사항을 반영하는 등 구체적인 방법론적 이슈들에 대해 생각할 시간을 가질 수 있기 때문이다(Kim, 2011). 많은 시간과 자원이 요구되는 조사를 할 경우에 사전연구 수행이 특히 필요한 이유이다.

본 연구는 두 가지 목적을 가지고 있다. 첫째, 연구방법론 차원에서 사전연구에 대해 소개하고자 한다. 국내에는 아직 사전연구를 논의하는 내용을 담은 학술지를 찾아보기 어렵다. 몇몇 서적을 통해 질적 연구자의 실패담과 질적 연구 수행의 어려움을 나누고 있지만(김진희·김영순·김지영, 2015; 이현서·박선웅, 2018), 구체적으로 사전연구를 소개하면서 연구 과정상의 이슈들을 살펴보는 저술은 미흡한 편이다. 본 연구의 두 번째 목적은 연구자가 사전연구를 수행하면서 깨닫게 된 질적 연구의 방법론적 이슈들과 사전연구 수행 후 주 연구에 반영하게 된 이슈들을 구체적으로 살펴보는 것이다. 이를 통해 사전연구의 유용성과 필요성을 보여 주고자 한다.

보통 학술지는 결과물 위주로 보고되기 때문에 본질적으로 결과를 산출하기 위한 것이 아닌 사전연구의 내용을 출판하는 경우는 드물다(Kim, 2011; Watson et al., 2007). 그러나 사전연구 수행의 경험을 발표하는 것은 결과에 상관없이 중요하다(Arain et al., 2010). 많은 초보 연구자들이 어떻게 질적 연구를 시작해야 하는지 잘 모르는 채 시행착오를 거치거나, 연구가 제대로 진행되는지 불안해하며 좌충우돌 연구를 진행한다(이현서·박선웅, 2018). 사전연구 노력의 결과를 학술 커뮤니티와 나누는 것은 불필요한 자원과 노력의 반복을 방지하며, 동료 연구자들이 윤리적이고 과학적인 연구를 수행할 수 있도록 돕는 연구자의 의무이다(Thabane et al., 2010).

이에 본 연구자는 연구자의 사명을 띄고 다음과 같은 내용을 살펴보고자 한다. 첫째, 사전연구의 정의 및 유용성에 관한 여러 학자들의 의견을 살펴볼 것이다. 둘째, 연구자의 배경과 결혼이민여성에 관한 연구 주제를 소개하며, 어떤 이유로 사전연구를 하게 되었는지에 관한 내용을 정리하고자 한다. 셋째, 사전연구를 수행하면서 깨닫게 된 질적 연구의 방법론적인 이슈들을 구체적으로 논의할 것이다. 마지막으로, 초보 연구자의 실제 경험을 바탕으로 질적 연구를 수행하는 데 있어 사전연구의 유용성 및 가치를 강조하며 결론을 맺고자 한다.

II. 사전연구의 개념과 유용성

네이버 지식백과에 ‘pilot study’라는 용어를 검색하면, 사전연구, 예비연구, 탐색연구, 타당성 조사 등 다양한 용어로 번역되고 있는 것을 볼 수 있다. 본 연구에서는 사전연구라는 용어로 통일하여 사용하고자 한다. 사전연구는 주 연구를 준비하는 차원에서 실시하는 소규모의 방법론적 실험이며, 연구자가 관심을 갖고 있는 주제가 현실적으로 수행 가능한 연구 질문이 될 수 있는지 파악하기 위해 시행하는 예비적 차원의 연구이다(Jariath et al., 2000; Kim, 2011; Prescott and Soeken, 1989; van Teijlingen and Hundley, 2002). 다시 정리하면, 사전연구는 연구 설계부터 현장연구 절차, 자료수집 계획과 도구 점검, 그리고 자료 분석 계획에 이르기까지 연구의 전 과정을 정교하게 만드는 데 도움을 주는 것으로 본 연구(main study)를 위한 연습의 일환이라고 할 수 있다(Kim, 2011; Yin, 2011).

비록 사전연구라는 용어가 타당성 조사(feasibility study)와 혼용하여 사용되는 경향이 있지만(Arain et al., 2010; LaGasse, 2013), 두 용어의 의미는 다소 차이가 있다. 사전연구나 타당성 조사 모두 본 조사(main study) 이전에 진행된다는 점에서 공통적이다. 그러나 차이점은 다음과 같다. 사전연구는 주 연구(main study)의 작은 버전이라 할 수 있다. 사전연구라는 용어를 사용할 때에는 연구의 모든 구성요소, 즉 연구자부터 조사 도구, 연구 참여자에 이르기까지 모든 요소들이 제대로 작동하는지 본 조사를 진행하기 전에 점검한다는 의미가 강하다(Arnold et al., 2009; Arain

et al., 2010; Leon et al., 2011). 반면, 타당성 조사는 주로 조사 대상자 모집 방법과 조사도구 등의 실행가능성을 점검하거나, 본 조사의 대략적 규모를 결정하기 위해 시행한다는 의미가 있다(Leon et al., 2011; Morin, 2013).

서론에서 언급한 것처럼, 질적 연구의 특성에 따라 굳이 사전연구를 하지 않아도 된다고 보는 입장(Holloway, 2008), 사전연구 수행이 오히려 연구자의 깊은 탐색에 방해가 될 수 있다는 입장(Morse, 1997) 등 사전연구 수행에 대한 반대의 입장이 존재한다. 그러나 대부분 질적 연구자들은 그 유용성에 훨씬 큰 비중을 두며 사전연구의 필요성을 주장하고 있다.

사전연구 경험을 들려준 서구의 질적 연구자들에 의해 사전연구가 어떤 점에서 어떻게 구체적으로 유용하다는 것인지를 살펴본 결과, 다음과 같이 대략 여섯 가지로 정리할 수 있었다. 첫째, 사전연구를 실시함으로써 연구자는 주 연구(main study)를 본격적으로 시작하기 전에 어떤 부분을 어떻게 조정하거나 변경해야 하는지 생각해 볼 수 있다(Kim, 2011; Wray, Archibong and Walton, 2017). 둘째, 연구 질문과 연구자가 염두에 두고 있는 연구 방법이 과연 적합한지 그리고 얼마나 실행 가능한 것인지 탐구하는 기회를 가질 수 있다(Kim, 2011). 셋째, 사전연구를 수행하면서 자료수집과 분석에 필요한 인적·물적 자원의 규모를 어느 정도 파악할 수 있다(Doody and Doody, 2015). 넷째, 두 명 이상의 연구자가 공동으로 연구를 수행할 때에, 사전연구를 통해 연구계획에 대한 전체 팀의 이해가 증진될 수 있다(Morrison, Clement, Nestel and Brown, 2016). 다섯째, 사전연구를 수행하면서 연구자는 본 연구가 시작되기 전부터 자기반영(self-reflection)하는 연습을 하게 되고, 이를 통해 연구를 보다 투명하며 윤리적으로 진행할 수 있게 된다(Doody and Doody, 2015; Kim, 2011; Morrison, Clement, Nestel and Brown, 2016). 여섯째, 명확한 목적을 가지고 사전연구를 수행할 때에 질적 연구의 엄격성과 신뢰성이 향상되며, 질 높은 연구를 수행할 가능성이 높아진다(Kim, 2011).

특히 초보 연구자의 경우 사전연구를 수행하면서 얻는 이점이 더욱 뚜렷하다고 한다(Doody and Doody, 2015; Wray et al., 2017). 현상학적 방법을 활용하여 학습 장애를 가진 노인을 돌보는 간호사의 경험을 탐구한 Doody와 Doody(2015)는 사전연구를 하면서 심층면담 하는 기술을 향상시킬 수 있었다고 하였다. 지식형성의 차원

에서 연구 참여자의 내부자적 경험을 탐색하는 면담이 보통의 의사소통 또는 치료적 면담과 어떻게 비슷하면서도 다른지에 대해 연습할 수 있었다는 것이다. Wray(2017)는 간호사의 문화적 역량의 실천 경험이라는 주제로 박사학위 논문을 쓰면서, Doody와 Doody(2015)처럼 사전연구에서 진행한 면담 녹취록을 리뷰하며 면담 기술이 향상되었다고 하였다. 무엇보다 초보 연구자가 사전연구를 꼭 해야 하는 필요성으로서 연구 방법론에 대한 이해가 높아지고, 연구자로서의 자신감이 생기는 점을 꼽고 있었다(Wray, 2017).

III. 연구자 배경 및 연구 주제에 대한 소개

본 연구자는 다문화 사회복지를 공부하는 중국인 유학생으로서, 특히 결혼이민여성의 사회참여 및 인적자본 개발에 관심을 갖고 있다. 유학생으로 한국에서 오 년째 살면서 한국 사회가 타 문화권의 사람을 어떻게 대하는지, 반대로 타 문화권의 사람이 한국 사회를 어떻게 인식하는지에 대해 경험하고 있는 덕분에, 연구자는 자연스럽게 문화적인 민감성을 갖추게 되었다고 생각하고 있다. 연구자는 중국의 대학에서 한국어 전공 학부과정으로 4년 동안 한국어를 배웠다. 2014년에 한국의 대학원에 입학하여 사회복지 전공으로 석사학위를 받았으며, 2019년 2월 현재 박사과정 재학 중에 있다. 연구자는 한국에서 유학하며 사회복지관련 이론적 지식을 습득하는 동시에 다양한 실천적 경험을 쌓으며 연구자로서 성장하기 위해 노력하고 있는 초보 연구자이다.

본 연구자는 이론적 지식을 현장에서의 실천으로 연결시키기 위해 다문화가족지원센터, 이주여성인권센터, 종합사회복지관 등에서 자원봉사 활동을 하면서 결혼이민여성들을 만나게 되었다. 이들과 함께 통·번역활동, 이주여성인권캠페인 등 다양한 사회참여 활동을 하면서, 결혼이민여성의 삶에 대해 점점 관심을 가지게 되었다. 이러한 경험을 바탕으로 연구자는 한국 사회에 결혼이민여성에 대한 여러 가지 인식이 실재하는 것을 보다 구체적으로 체험하게 되었다. 보통 한국 사람들의 결혼이민여성에 대한 인식은 가난한 나라에서 더 나은 삶을 찾아 한국사회로 이주한 사람들, 한국 사회에의 적응이 필요한 복지수혜자 정도에 머물고 있다(신현태·정우열·유근환, 2012; 정복

동, 2016). 그러나 본 연구자가 사회참여 활동을 하면서 만난 결혼이민여성들은 자기 삶의 주체가 되고 싶어 하며, 자기 모국어와 한국어를 동시에 구사하고, 기꺼이 문화적 역량을 높이고 싶어 하는 인격체들이었다.

본 연구자의 석사학위 논문은 중국 결혼이민여성의 자원봉사에 관한 것이었다. 사회적응과 재능기부 차원에서 다양한 자원봉사 활동에 참여하고 있는 중국 결혼이민여성들은 자원봉사 활동에 만족하지만 직업을 갖고 보다 주체적인 삶을 원하고 있었다. 이들은 중국과 한국의 이중 문화에 능통한 자신의 능력을 살리고, 고등교육 및 전문직 훈련을 통해 전문가로 자리매김하기를 원하고 있었다. 그리고 한국의 사회복지유관기관들에서는 이들을 통·번역, 상담 업무 등을 통해 새로 정착하는 결혼이민여성들을 돕는 준전문가로 활용하는 것을 볼 수 있었다(김유진·유전양, 2018).

이런 경험을 바탕으로, 본 연구자는 결혼이민여성의 역량강화를 구체적으로 지원할 수 있는 방법에 대해 고민하게 되었다. 한국 사회는 동화주의적 사회적응을 강조하던 분위기에서 점차 다양한 문화적·언어적 자원을 인정하고 활용하는 분위기로 변해가고 있다(김유진·유전양, 2018). 이러한 시대적 흐름 가운데 결혼이민여성의 한국 내 정착이 많아지면서 이들에 관한 연구도 계속 증가하고 있다. 하지만 결혼이민여성의 사회통합을 위한 논의 및 취업 활동에 관한 연구는 아직 미흡한 편이다(이정희·이수분, 2013). 주로 결혼이민여성의 취업현황 조사, 취업 동기 및 취업만족도에 관한 요인 검증 위주의 연구가 진행되었고(이현주·여영훈, 2017; 이혜림·조민효, 2014), 이들의 사회적·경제적 자립 또는 역량강화에 관한 심층 연구는 찾아보기 어렵다.

결혼이민여성들의 사회에 기여하고 싶어 하는 마음, 단순히 이민사회에 적응하는 차원이 아니라 능동적인 사회 구성원으로 참여하고 싶은 마음을 어떻게 지원할 수 있는가? 많은 결혼이민여성들이 동료, 후배 결혼이민여성을 더 잘 돕기 위해 상담이나 사회복지를 공부하는데, 이들이 준전문가 및 전문가로 성장할 수 있도록 어떻게 도울 수 있는가? 이런 질문에 답할 수 있는 관련 연구가 현재 충분하지 않기 때문에, 연구자는 먼저 현장에 들어가 부딪히며 배워야겠다고 생각하게 되었다. 이에 따라, 본 연구자는 박사과정에 진학하면서 결혼이민여성이 준전문가 또는 전문가로 성장하는 경험에 대해 연구하고자 하였다. 그런데 구체적으로 연구 질문을 어떻게 구성해야 할지 막막했다. 앞서서도 언급했듯이, 준전문가와 관련한 선행연구를 찾기가 어려웠기 때문이

다. 어디서부터 출발해야 하는지 고민하며 석사학위논문의 연구 참여자들을 다시 떠올리게 되었다. 열 명의 중국 출신 결혼이민여성들 중에서 과반수 이상이 사회복지사 또는 상담가가 되기를 희망하며 대학에 진학하거나 진학 계획을 세우고 있던 것이 생각났다. 이런 생각의 흐름 속에서 결혼이민여성의 대학교육, 그중에서도 사회복지 전공 교육 관련 선행연구를 찾아보았다. 역시 관련 선행연구와 이론적 틀을 찾기가 쉽지 않았다. 대학에 재학 중인 결혼이민여성을 직접 만나 그들의 이야기를 들어볼 수밖에 다른 방법이 없었다. 이 주제에 관한 현상과 경험을 살펴보기 위해 일단 대학에서 사회복지를 전공하는 결혼이민여성을 편의표집 방법으로 접촉하게 되었다. 사전연구는 이렇게 시작되었다.

사전연구 수행을 위해 경북대학교 생명윤리심의위원회 심의(IRB)를 통과하고 본격적으로 연구 참여자들을 접촉하기 시작하였다. 연구자의 석사학위논문의 연구 참여자와 형성된 관계를 통해서 눈덩이 표집 방법으로 모집을 시도하였고, 사회복지 전공으로 대학에 재학 중인 여섯 명의 결혼이민여성을 만날 수 있었다. 2017년 12월에 ○○대학교 다문화복지한국어학과와 3학년 학생들이 겨울방학 동안 사회복지현장실습을 수강하고 있었고, 그에 맞추어 사전연구 자료 수집을 시작하였다. 모두 여섯 명의 연구 참여자와 인터뷰를 진행하며 각 연구자의 녹취내용을 정리하였다. 연구자는 각 연구 참여자로부터 연구목적 설명서 및 연구 참여 동의서에 승인을 받았고, 연구 참여자가 원하는 시간과 장소에서 심층면담을 진행하였다. 중국 출신 연구 참여자의 경우는 중국어로 인터뷰를 진행하였고 일본과 베트남 출신의 연구 참여자의 경우는 한국어로 진행하였다. 각각의 인터뷰는 거의 두 시간 정도의 시간이 소요되었고 인터뷰의 전체 내용은 모두 녹취하였다. 매번 인터뷰를 마칠 때마다, 연구자는 연구 참여자에게 인터뷰 내용 중 모호한 부분을 다시 확인하고 수정하는 작업을 진행하였다(Lincoln and Guba, 1985). 참여관찰을 포함하여 연구 진행 과정에서 일어난 일, 지도교수님과의 연구 관련 대화, 선행연구를 읽으면서 연구와 관련하여 연구자가 생각하게 된 내용 등을 연구노트에 기록하였다.

IV. 사전연구를 통해 드러난 연구 수행의 어려움과 방법론적 이슈들

1. 연구 도구로서의 자기반성

질적 연구에서 연구자는 자료수집과 분석의 주요 도구이다(Padgett, 1998; Yin, 2014). 모든 연구자는 세상을 바라보는 방식, 삶의 방식에 대한 생각, 자신의 개인적·학술적 배경, 그리고 해당 연구 주제에 대한 이론적 틀의 영향을 받는다. 알게 모르게 이러한 영향을 받고 있기에, 연구자마다 어떤 것에 초점을 두며 참여관찰을 하는지, 어떤 것에 초점을 맞추며 면담을 하는지, 그리고 줄 단위 코딩 후에 의미단위를 어떻게 묶어나가며 분석하는지 등이 달라진다. 연구 시작부터 마침까지 전반에 걸쳐 연구자라는 필터를 거치지 않는 부분이 없는 것이다. 문자 그대로, 질적 연구에서 연구자는 연구의 주요 도구이다(use of self). 연구자의 정체성과 세계관, 전문적 지식의 수준과 깊이, 융통성과 통찰력, 의사소통기술 등에 따라 수집되는 자료의 내용과 수준이 달라지고, 분석의 결과도 달라지게 된다.

본 연구자는 질적 연구 방법론 수업을 수강하였고, 질적 연구 방법으로 석사학위논문을 작성하였지만, 여전히 질적 연구 방법이 어렵고, 연구자로서 역량이 매우 부족하다고 생각하였다. 연구 도구로서 갖추어야 할 역량인 자기반영성(self-reflection), 창의성, 유연성, 이론적 민감성(theoretical sensitivity), 추상적이고 개념적으로 사고하며 새로운 개념을 만드는 기술, 그리고 명료하면서 깊이 있는 글쓰기, 모든 부분에서 자신이 없다. 다행히 본 연구자는 개방적이며, 의사소통을 잘하고, 사람 대하는 것을 어려워하지 않는다. 그런데 연구자의 문화적 역량에 대해 곰곰이 생각하면서, 다른 문화권의 사람들을 사교적인 만남이 아닌 연구를 위해 만난다는 것이 부담으로 다가왔다. 석사학위논문을 위한 연구에서는 같은 문화권의 참여자들을 만났지만, 이번 연구에서는 중국 이외의 베트남, 일본 결혼이민여성들과도 심층면담을 하기 때문이다.

연구자는 사전연구를 수행하면서 연구 도구로서 자신의 활용에 대해 연습하는 기회로 삼기로 했다. 즉, 사전연구를 연구자로서의 자기반성(self-reflection)의 시간으로

삼기로 했다. 사회복지학자이자 질적 연구자인 Padgett(1998)은 연구자가 자신을 돌아보며 자신의 신념체계와 편견, 관계 맺는 방식 등을 인식하고, 이러한 것들이 연구에 어떤 영향을 미치는지 이해하기 위해 노력해야 한다고 하였다.

어느 날 결혼이민여성들과 면담을 진행하면서 ‘참여자는 나를 어떻게 볼까?’하는 생각을 하게 되었다. ‘나는 도대체 여기에 연구자로 있는 것인가?’ 아니면 ‘연구 참여자에게 동생뻘의 사람처럼 비춰지는가?’라는 의구심이 들면서 연구자의 역할, 그리고 연구자가 연구 참여자에게 어떻게 다가가야 하는가에 대해 고민하게 되었다. 여섯 명 사전연구 참여자의 평균 연령은 42.3세로 가장 나이가 적은 베트남 출신의 결혼이민여성이 34세였다. 27세의 연구자는 연상의 결혼이민여성들에게 자연스럽게 ‘누나’라고 호칭하며 관계형성(rapport)을 한다고 생각했었다. 중국에서는 한국과 마찬가지로 혈연관계가 아니라도 연상의 여자를 부를 때 ‘지에게(姐姐)’라는 호칭을 사용한다. 일본과 베트남 등 다른 아시아권에서도 그러한지는 미처 알지 못했지만, 연구자는 연구 참여자들에게 자연스럽게 ‘누나’라고 부르곤 하였다.

면담의 분위기를 되새겨 보며 그리고 녹취록을 읽으면서, ‘누나’라는 호칭이 연구자의 의도만큼 자연스럽게 않을 수도 있다는 생각을 하게 되었다. 어떤 참여자는 연구자를 편하게 대하며 자신의 이야기를 들려주었지만, 또 다른 참여자는 인터뷰를 사적인 대화로 여기는 듯한 느낌을 갖게 되었기 때문이다. 연구자가 어떤 이유로든 연구자의 역할을 제대로 참여자에게 인식시켜 주지 않으면 인터뷰 내용의 진실성이 떨어지고 연구의 신뢰성에 대한 위기가 생길 수 있다(Kim, 2011). 이러한 문제를 방지하기 위해서 연구자와 연구 참여자 간에 어느 정도의 라포가 형성되면 연구자는 자신의 역할에 초점을 맞추며 참여자 역시 연구에 참여하는 역할에 집중할 수 있게 도울 수 있어야 한다. 이와 관련하여 Padgett(1998)은 질적 연구자는 관계형성에 신경을 쓰면서 객관적인 거리감을 유지하고, 연구 참여자가 연구 질문에 초점을 맞추어 의사소통할 수 있도록 도와야 한다고 하였다.

연구자로서 자신의 편견과 이것이 연구 참여자와의 상호작용에 미치는 영향에 대해서 생각하게 된 다른 이슈도 있다. 바로 연구자와 연구 참여자의 민족적 배경의 차이에 대해 연구자가 지나치게 의식한다는 점이었다. 아무래도 문화적 민감성을 중요하게 여기는 다문화 복지를 연구하며, 연구자의 인종 및 민족성과 참여자의 인종 및 민족성

이 일치하는 관계에서 참여자의 신뢰가 더 높다는 연구 방법론 교재의 내용을 의미심장하게 받아들였기 때문인 것 같았다. 중국인 결혼이민여성 참여자를 만날 때와 달리, 베트남 또는 일본 결혼이민여성과의 면담을 앞두고 평소와 다르게 긴장을 하거나 초조함을 느끼기도 하였다. 방법론 교재를 계속 공부하면서 인종 및 민족성이 다른 관계의 장점(예, 연구자와 연구 참여자 모두 새로운 시각을 갖게 되는 기회가 될 수 있다)과 민족성의 매치보다 더 중요한 것은 연구 참여자와의 관계형성을 위한 연구자의 노력이라는 것을 깨닫고 난 후에는 더 이상 긴장하지 않게 되었다. 관계형성을 위한 노력은 연구자로서 문화적 역량을 키우는 것, 그리고 융통성을 갖추는 것과는 관련 있다고 생각하게 되었다.

Padgett(1998)은 질적 연구자는 높은 수준의 언어적·문화적 역량을 갖추어야 한다고 하였다. 본 연구에서 연구자는 중국인이고 연구 참여자는 중국, 일본, 베트남 결혼이민여성들이다. 중국 출신 연구 참여자의 경우는 중국어로 인터뷰를 진행하고 일본과 베트남 출신의 연구 참여자의 경우에는 한국어로 진행하였다. 그 후에 연구자가 녹취한 내용을 한국어로 풀고 각 사례의 연구일지를 작성하였다. 중국인 연구 참여자와의 인터뷰는 모국어인 중국어로 유창하게 소통할 수 있는 장점이 있는 반면에 중국어로 녹취한 인터뷰 내용을 문서형식으로 풀 때 연구자가 한국어를 번역하는 능력에 따라서 의미 전달이 제한되는 것을 볼 수 있었다. 그리고 일본과 베트남 출신의 연구 참여자와 같이 한국어로 진행한 인터뷰의 경우에는 연구자와 연구 참여자가 모두 모국어가 아닌 한국어를 사용함에 따라 한국어 능력의 제한점으로 인해 각 참여자가 전달하고 싶은 내용과 차이가 나는 것을 느낄 수 있었다.

연구자는 언어적인 능력뿐만 아니라 타 문화권의 습관, 전통, 풍습 등을 미리 충분히 이해하고 있어야 연구 참여자와의 관계형성이 어렵지 않을 것이다. 면담 녹취록을 되풀이해서 청취하면서, 연구자가 연구 참여자의 신념, 가치관, 사회문화적 배경 등에 개방적이고 포용적인 가치관을 지니는 것이 매우 중요하다는 것을 알게 되었다. 연구자는 자신의 견해, 가치관, 신앙 등이 다르거나 심지어 적대적인 연구 참여자를 만날 수도 있다. 이때 연구자가 폐쇄적이고 보수적인 가치관을 갖고 있으면 이러한 태도가 연구 참여자에게 투영될 수 있을 것이다. 연구자는 어떤 경우에도 연구 참여자를 존중하며, 이를 연구자가 알 수 있도록 해야 한다고 생각하게 되었다.

2. 연구 주제 변경에 따른 연구 설계 재검토

면담을 하면서 결혼이민여성들이 한국 정부와 지방자치단체의 지원으로 대학에 진학한 것을 볼 수 있었다. 연구 참여자가 구술한 내용의 상당 부분이 주로 입학조건, 수업진행절차, 실습기관의 협조, 자격증 취득에 대한 조건 및 앞으로 취업에 대한 고민 등에 대한 것이었다.

현재 연구 참여자들은 ○○대학교 다문화복지한국어학과에 재학 중인데, 이 학과는 2015년 전국 최초의 다문화특화사업으로 결혼이민여성 대상의 대학교육 과정이다. 결혼이민여성들의 대학교육 경험은 지방자치단체의 다문화사회통합정책과 영향을 주고받는 맥락 내에서 진행 중인 것이다. 연구자는 연구 참여자와 인터뷰를 진행하면서 이 사업에 관련된 기관의 업무와 수행기능에 대해서 알게 되었고, 사업 이해관계자 간의 상호 협력적 관계가 결혼이민여성의 전문적 성장과 더불어 취업에 핵심적인 영향을 미치는 것을 알 수 있었다. 거시적인 시각으로 보면, 한국 다문화사회통합의 과정에서 ○○대학교의 지방자치단체 다문화 특화사업사례는 많은 유관기관(시청, 대학, 다문화가족지원센터, 사회복지실습기관, 자격인증기관, 지방고용노동청)의 상호 협력을 통해 결혼이민여성의 사회참여가 이루어지는 과정을 보여준다. 그리고 미시적인 시각으로 보면, 결혼이민여성들의 준전문가에 대한 역량강화과정은 자신의 문화적 요소, 입학 동기, 이론적 지식을 습득한 경험, 실습경험, 취업에 대한 기대 등 많은 개인적 요소들이 있음을 알 수 있다.

결혼이민여성의 경험에만 초점을 두어 질문하면 연구 결과가 한정적일 수밖에 없다고 생각하게 되었다. '다문화가족 맞춤형 대학지원 사업'이 구체적으로 어떤 사업인지에 대해 알아보고 연구주제를 보다 거시적인 시각으로 바라볼 필요성이 있음을 자각하게 하였다. 이러한 문제 인식을 통해 주 연구의 연구 질문으로서 "다문화가족 맞춤형 대학지원 사업을 통해 살펴보는 대한민국 결혼이민여성들의 대학교육과 취업의 현실"로 재설정하기로 하였다. ○○대학교의 시범적인 사업이 어떤 과정을 통해 정책으로 실행되었고, 사업의 성과를 어떤 목적과 목표로 달성하고자 하며, 그 가운데 결혼이민여성을 비롯한 이해관계자들의 경험은 어떠한지를 연결시켜 살펴보기로 하였다. 이와 같은 연구 질문을 탐색할 수 있는 최선의 방법론으로서 Stake식의 도구적 사례연구

방법을 생각하게 되었다. 앞에서 정리했듯이, 원래 사전연구를 실행했던 이유는 결혼 이민여성이 사회복지사로 성장하는 준전문가로서의 과정에 대한 여성의 경험에 대해 이해하는 것이었다. 그리고 질적 기술연구(qualitative descriptive method) 방법을 활용할 계획이었다.

질적 연구 방법론 교재에서는 연구 방법론과 연구 질문 간의 관계를 마치 닭과 달걀의 관계처럼 설명하고 있다. 연구 질문이 먼저 정해져야 어떤 방법으로 어떻게 연구를 할 것인가의 방향이 자연스레 정해진다는 입장이 있다. 한편, 먼저 자료 수집을 하면서 현장 경험을 하다 보면, 자연스레 연구 질문이 구체화된다고 바라보는 입장도 있다(Yin, 2011). 초보 연구자는 이것이 무슨 말인지 직접 경험하기 전에는 알 수 없다. 본 연구자의 경우에는 사전연구를 수행하면서 연구 질문이 바뀌게 되었다. 그에 따라 방법론도 바뀔 수 밖에 없게 되었다. 처음에 설계한 대로 연구가 진행되지 않는다고 해서 당황할 필요는 없다. 연구 설계는 연구 문제, 자료 수집과 분석 방법을 연결시켜서 연구 결과가 원래 의도했던 연구 문제를 다루도록 하는 논리적 청사진이라고 할 수 있다(Yin, 2011). 양적 연구방법과 달리 질적 연구에서는 연구가 진행되면서 설계가 바뀔 수 있다(Yin, 2011). 즉, 수정된 연구 질문에 답할 수 있는 방법론이 처음에 생각했던 것이 아닐 수 있다는 것이다. 다만 초보 연구자로서 부담되는 것은 연구 질문에 따라 가장 최선의 방법론이 무엇인지 제대로 맞추기가 쉽지 않다는 것이다. 최선의 방법론을 선택할 수 있을 정도로 방법론에 대한 이해가 필요하다는 것을 의미한다.

3. 심층 면담 질문 추가 및 변경

연구 질문이 바뀌었기 때문에, 본 연구(main study)가 시작되면 자료 수집의 방법과 범위, 면담 질문도 변경할 예정이다. 먼저, 연구 참여자의 경험을 이해하기 위해 기존에 준비했던 질문 목록에 결혼이민여성 맞춤형 대학지원사업의 관계자와 연구 참여자 간의 상호작용에 관한 질문을 추가할 것이다. 둘째, 사회복지현장 실습에 초점을 두었던 질문을 사회복지현장실습과 대학의 교육 전 과정에 대한 질문으로 확대하여 질문할 것이다. 셋째, 향후 주 연구에서는 연구 참여자를 결혼이민여성뿐만 아니라 사업 관계

자들(대학, 시청, 다문화가족지원센터 등)로 확대하여 면담을 진행할 것이다.

V. 결론

연구자의 지도 교수님은 연구자의 렌즈에 대해 종종 말씀하신다. 빨간 색 렌즈로 세상을 보면 세상이 빨강계 보이고, 파란 색 렌즈로 세상을 보면 세상이 파랗게 보인다는 말씀을 하시며, 개념적 틀의 중요성에 대해 강조한다. 앞에서 기술했듯이, 연구자는 연구의 주요 도구이고(use of self), 연구자라는 필터를 거쳐 연구 결과와 결론이 도출된다. 연구자가 하는 모든 활동은 연구자의 세상을 바라보는 방식, 얹의 방식에 대한 생각, 자신의 개인적·학술적 배경, 그리고 해당 연구 주제에 대한 이론적 틀의 영향을 받는다. 연구자가 자신의 렌즈가 어떤 색상인지 제대로 인식하고 지속적으로 자기 성찰을 하는 것이 필요한 이유이다. 본 연구자에게는 사전연구 수행 경험을 나누는 작업이 바로 그러한 역할을 하고 있다.

연구자는 결혼이민여성의 역량강화에 깊은 관심을 갖고 있지만, 선행연구가 미흡한 상황에서 이를 어떻게 연구로 구체화할지 알 수 없었다. 일단 연구 현장에 들어가서 부딪히면서 배우기로 하였다. 그리고 사전연구를 수행하며 연구 도구로서 자신을 성찰하고, 총체적 연구 능력을 점검하였다. 연구 도구로서 연구자 자신의 역량이 연구 참여자 및 연구결과에 미치는 다양한 영향을 깨닫고, 질적 연구자로서 융통성과 문화적·언어적 역량이 중요한 자질임을 알게 되었다. 본 연구자가 보기에, 사전연구는 초보 연구자에게 연구자로서 자기계발하는 기회이다. 사전연구를 통해 연구자는 참여 관찰 기술, 심층면접 기술 뿐 아니라 무엇보다 연구자로서 자신의 자질과 특성, 부족한 부분과 잘 하는 부분을 돌아보게 되며, 한층 성숙한 연구자로 성장하는 시간을 갖게 된다. 그리고 사전연구를 통해 연구 주제를 재점검하는 기회를 가질 수 있었다. 연구 주제를 검토하면서 초보 연구자에게 막연할 수 있는 연구의 설계에 대해서 고민하며, 질적 연구에 대한 이해가 깊어지는 시간을 갖게 되었다. 연구자가 설정한 초기의 연구방법(일반적인 질적 기술연구)을 향후 주 연구에서는 도구적 사례연구 방법으로 변경하기로 한 것이다. 그리고 사전연구의 경험을 바탕으로, 면담 질문이 수정·추가되

고, 연구 참여자의 범위도 다문화가족 대학교육시범사업의 관계자까지 확대하기로 하였다.

사전연구를 통해 주 연구를 시행할 때 발생할 수 있는 다양한 문제들, 즉 연구주제, 연구대상과 모집방법, 연구방법 등을 조정하고 수정할 수 있는 기회를 가짐으로써, 주 연구과정에서 발생할 수 있는 시행착오를 줄이고 초보 연구자로서 역량을 강화할 수 있다는 점에서 사전연구의 의의가 있다고 할 수 있겠다. 본 논문에서는 질적 연구의 경험을 바탕으로 사전연구의 유용성과 필요성에 대해 살펴보았지만, 사전연구 수행은 양적이든 질적이든 연구 질문에 따라 연구하는 과정에서 부딪히게 되는 다양한 고민에 대해 새로운 시각을 제공할 수 있다. 그리고 실제 주 연구를 진행하기에 앞서 관련 요소들을 파악할 수 있도록 한다. 초보 연구자는 이러한 연구과정을 통해 연구 주제에 관한 사고의 폭이 확장되면서 연구자 자신에 대한 인식도 새롭게 변화하게 된다.

이번 사전연구 수행을 통해 무엇보다 문헌고찰의 중요성과 방법론 공부의 필요성을 체감하게 되었다. 소위 ‘아는 만큼 보인다’고 하는데, 막상 연구를 시작하였지만, 연구자가 무엇을 알고 어떤 것은 모르는지조차 알 수 없었다. google 학술검색 페이지에 나오는 아이작 뉴턴의 문장, ‘거인의 어깨에 올라서서 더 넓은 세상을 바라보라’에 대해서도 곰곰이 생각하게 되었다. 자료에 의미를 부여하고 중요한 것과 중요하지 않은 것을 구별하며, 연구 주제가 되는 현상을 의미 있게 그려낼 수 있는 능력을 갖추는 것, 앞으로 집중해서 키워가야 할 연구자의 역량이다.

참고문헌

- 김유진·유전양. 2018. “중국 결혼이민여성의 참여경험을 바탕으로 살펴본 국내 다문화 자원봉사 사업의 명암(明暗).” 『사회과학연구』 제29권 1호: 239-257.
- 김인숙. 2016. 『사회복지연구에서 질적방법과 분석』. 서울: 집문당.
- 김진희·김영순·김지영. 2015. 『질적연구 여행』. 성남: 북코리아.
- 신현태·정우열, 유근환. 2012. “다문화사회와 사회통합에 관한 탐색적 연구-주요국과 한국의 다문화정책 비교를 중심으로.” 『한국지방자치연구』 제13권 4호: 177-200.
- 윤택림. 2013. 『문화와 역사 연구를 위한 질적연구 방법론』. 서울: 아르케.
- 이정희·이수분. 2013. “여성결혼이민자의 직업적응과정에 관한 질적연구.” 『인적자원관리연구』 제20권 1호: 141-159.
- 이현서·박선용. 2018. 『질적 연구자 좌충우돌기』. 파주: 한울엠플러스.
- 이현주·여영훈. 2017. “결혼이주여성 근로자의 직무적, 문화적 요인이 직무만족에 미치는 영향.” 『예술인문사회융합멀티미디어논문지』 7호: 415-423.
- 이혜림·조민효. 2014. “다문화가정 이주여성의 적응유형과 사회자본이 취업에 미치는 영향에 관한 연구.” 『한국정책학회보』 제23권 4호: 237-263.
- 정복동. 2016. “한국의 다문화정책 변화와 과제-다문화가족서비스 전달체계를 중심으로.” 『한국사회공헌학회지』 3호: 21-38.
- Akerlind, G. 2005. “Learning about phenomenography: Interviewing, data analysis, and the qualitative research program.” *Doing Developmental Phenomenography*. RMIT University Press, Australia 63-73.
- Arain, M., Campbell, M. J., Cooper, C. L., and Lancaster, G. A. 2010. “What is a pilot or feasibility study? A review of current practice and editorial policy.” *BMC medical research methodology* 10(67): 1-4.

- Arnold, D. M., Burns, K. E., Adhikari, N. K., Kho, M. E., Meade, M. O., and Cook, D. J. 2009. "The design and interpretation of pilot trials in clinical research in critical care." *Critical care medicine* 37(1): S69-S74.
- Chiswick, B. R. and DebBurman, N. 2004. "Educational attainment: analysis by immigrant generation." *Economics of Education Review* 23(4): 361-379.
- Clement, T., Brown, J., Morrison, J., and Nestel, D. 2016. "Ad hoc supervision of general practice registrars as a 'community of practice' : Analysis, interpretation and re-presentation." *Advances in health sciences education* 21(2): 415-437.
- Creswell, John W. 2013. *Qualitative and Research Design: Choosing Among Five Approaches, 3rd Edition*. SAGE Publications, Inc.
- Doody, O., and Doody, C. M. 2015. "Conducting a pilot study: Case study of a novice researcher." *British Journal of Nursing* 24(21): 1074-1078.
- Holloway, I. 2008. *AZ of qualitative research in nursing and healthcare*. Wiley-Blackwell, Chichester.
- van Teijlingen, Edwin R. and Hundley, Vanora. 2002. "The role of pilot studies in midwifery research." *RCM midwives: the official journal of the Royal College of Midwives* 5(11): 372-374.
- Jairath, N., Hogerney, M., and Parsons, C. 2000. "The role of the pilot study: A case illustration from cardiac nursing research." *Applied Nursing Research* 13(2): 92-96.
- Kezar, A. 2000. "The importance of pilot studies: Beginning the hermeneutic circle." *Research in Higher Education* 41(3): 385-400.
- Kim, Y. 2011. "The pilot study in qualitative inquiry: Identifying issues and learning lessons for culturally competent research." *Qualitative Social Work* 10(2): 190-206.

- LaGasse, A. B. 2013. "Pilot and Feasibility Studies: Application in Music Therapy Research." *Journal of Music Therapy* 50(4): 304-320.
- Leon, A. C., Davis, L. L., and Kraemer, H. C. 2011. "The role and interpretation of pilotstudies in clinical research." *Journal of psychiatric research* 45(5): 626-629.
- Lincoln, Yvonne and Guba, Egon. 1985. *Naturalistic Inquiry*. Beverly Hills, CA: SAGE.
- Marshall, C. and Rossman, G. 2006. "The how of the study: Building the research design." *Designing qualitative research* 55-101.
- Morin, K. H. 2013. "Value of a pilot study." *Journal of Nursing Education* 52(10): 547-548.
- Morrison, J., Clement, T., Nestel, D., and Brown, J. 2016. "Underdiscussed, underused and underreported: pilot work in team-based qualitative research." *Qualitative Research Journal* 16(4): 314-330.
- Morse, J. M. 1997. "The pertinence of pilot studies." *Qualitative Health Research* 7(3): 323-324.
- National Institute for Health Research. 2012. *Evaluation, Trials and Studies Glossary*. <http://tinyurl.com/ppcusje>(accessed 10 November 2015).
- Padgett, Deborah K. 1998. *Qualitative Methods in Social Work Research: Challenges and Rewards*. Sage Publications, Inc.
- Prescott, P. A., and Soeken, K. L. 1989. "The potential uses of pilot work." *Nursing Research* 38(1): 60-62.
- Stake, R. E. 2000. *Case studies*. Handbook of qualitative research 2: 435-454.
- Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L., Robson, R., Thabane, M., Giangregorio, L. and Goldsmith, C. 2010, "A tutorial on pilot studies: the what, why and how", *BMC Medical Research Methodology* 10(1): 1-10

- Watson, F. A. 2016. "Lessons learned on approaches to data collection and analysis from a pilot study." *Nurse Researcher (2014+)* 24(1): 32.
- Wray, J., Archibong, U. E., and Walton, S. 2017. "Why undertake a pilot in a qualitative PhD study? Lessons learned to promote success." *Nurse Researcher* 24(3): 31-35.
- Yin, Robert K. 2011. *Qualitative Research from Start to Finish*. Guilford Publications, Inc.

Experience of conducting qualitative research by a novice
researcher in social science:
Based on practicing the pilot study

Chuanyang Yu*·Yujin Kim**·Mikyung Seo***

ABSTRACT

The purpose of this study is to introduce a pilot study in the aspect of research methodology and to investigate the difficulties and the methodological issues in the progress of the qualitative research which was realized while carrying out the pilot study. The pilot study is a small-scale methodological experiment conducted to prepare a main study, and it is a preliminary study to determine whether a topic of interest to the researcher can be a realistic and feasible. Despite its usefulness, the pilot study has not been adequately discussed in the qualitative research methodology textbooks. Based on the actual experience of a novice researcher who is a Chinese student, the purpose of this study was to emphasize the usefulness and value of the pilot study in carrying out the qualitative research.

Keywords: Pilot study, Qualitative Research, Use of Self, Methodology

투고일: 2019.02.15. 심사일: 2019.03.11. 게재확정일: 2019.03.12.

* Ph.D. Candidate, Faculty of Social Welfare, Kyungpook National University.

** Corresponding Author, Associate Professor, Faculty of Social Welfare, Kyungpook National University

*** Ph.D. Candidate, Faculty of Social Welfare, Kyungpook National University.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.21>

【연구논문】

휴대폰 사용은 조현병 환자에게 지각된 사회지지를 반영하는 지표가 될 수 있나?

우 리 님* · 최 성 진**

논문요약

본 연구는 정신건강과 휴대폰 사용양상이 사회지지 지각에 미치는 영향을 살펴 보았다. 27명의 조현병 환자와 27명의 정상 성인이 자기 보고형 설문지를 작성하였다. 측정은 휴대폰 사용 질문지, 한국판 일반정신건강질문지, 그리고 사회지지 척도를 이용하였다. 자료 분석에는 집단간 차이를 검증하였고, 회귀 분석을 실시하였다. 그 결과, 조현병 환자는 정상 성인에 비해 정신건강 상태가 양호하지 못했고, 사회지지를 덜 지각하였으며, 휴대폰을 덜 사용하였다. 또한 정신건강 상태가 좋지 못함에도 불구하고, 휴대폰 사용의 조절효과를 통해 사회지지를 보다 많이 지각할 수 있었다. 휴대폰 사용의 의사소통 기능은 사회지지에 유의한 효과가 있었지만, 도구적 기능은 그렇지 못했다. 문자 메시지를 선호하는 사람은 음성 통화 선호보다 사회지지 효과가 컸다. 휴대폰 사용은 사회적 상호작용과 정서적 지지를 증가시키고, 심리사회 스트레스의 조절 역할을 했다. 그리고 의사소통 기능을 가지고 있는 문자 메시지는 도구적 기능으로서의 휴대폰 사용보다 사회지지의 지각을 높이는데 도움을 주었다.

주제어: 휴대폰 사용, 사회지지, 정신건강, 조현병

I. 서론

우리나라 휴대폰의 상용서비스가 시작된 1997년에는 682만명이 휴대폰을 사용했는데, 2017년 현재 6300만명이 휴대폰을 사용하고 있어(과학기술정보통신부, 2018)

* 경상대학교 심리학과, 박사과정생.

** 동명대학교 상담심리학과 교수, 교신저자.

무려 10배 가량 사용자가 증가한 것을 확인할 수 있다. 하지만 이와 같은 휴대폰 시장의 급속한 성장의 이면에 그 부작용도 만만치가 않다. 미래창조과학부와 한국정보화진흥원(과학기술정보통신부, 2015)이 만3세 이상 59세 이하 18,500명을 대상으로 가구방문 대인면접조사를 실시한 결과, 스마트폰중독 위험군 비율은 지속적으로 증가하고 있었다. 이외에 여러 연구에서 휴대폰의 과다한 사용으로 인한 부정적 영향을 주로 서술하고 있는데, 정신건강과 관련된 변인과 휴대폰의 관련성을 설명하고 있는 연구가 많다. 예를 들어, 휴대폰의 중독적 사용군에서 신체 통증, 우울, 불안, 부정적 자기개념, 부정적 애착 수준, 충동성 등이 높았고(황경혜 외 2012; 장혜진 외 2006), 휴대폰의 중독적 사용이 많을수록 대인관계 문제, 스트레스, 수면문제, 우울증상이 많아지는 정적 상관을 보였다는(홍은영 2015; 토미 외 Thome'e, et al 2011) 결과도 있다. 하지만 대부분 청소년과 대학생을 대상으로 한 연구로, 정신건강과 직접적으로 관련된 정신장애인을 대상으로 한 휴대폰 관련 국내연구는 전무하다. 국외 연구에서도, 에인스워드 외(Ainsworth 2013)의 논문에서 정신장애인의 증상을 모니터링하는 방법으로써 휴대폰을 어떻게 사용하는 것이 효과적인가에 대한 연구를 실시했으나 이는 실무적 내용으로, 정신장애인의 휴대폰 사용과 심리적 특성의 관계를 살펴보지는 못했다.

한편, 정신장애인의 삶의 만족도에 영향을 미치는 주요 요인으로 사회지지가 있다(정명숙 외 2008). 사회지지는 관계 속에서 얻어지는 긍정적이고, 인지적이며, 도구적인 자원을 의미한다(코헨 Cohen 외 1983). 특히, 정보 지지나 물질 지지는 정신장애인에게 중요하고, 가족지지는 삶의 질을 향상시킨다(최일섭 외 2007). 또한, 빈곤과 같은 경제 스트레스나 외상 스트레스에서 나타나는 무력감과 죄책감, 그리고 지체부자유로 인한 사회적 어려움을 겪는 사람들에게도 사회지지는 심리사회 스트레스의 완충 역할을 하고, 긍정자원으로써 기능을 한다(엄태완 2008; 이지영 외 2008; 전현선 외 2008). 즉, 사회지지는 정신건강 관련 스트레스의 보호요인이 된다. 다른 측면에서, 정신장애인과 관련된 연구는 아니지만, SNS를 실행하고 그 안에서 적극적으로 커뮤니케이션 할수록 사회적 지지감이 증가한다는 결과(조윤경 외 2014)가 있어, 다양한 경로의 의사소통과 사회지지 간의 연결고리를 찾을 수 있다.

따라서 본래 사람들 간의 상호작용과 의사소통을 지지하기 위해 시작되었던 휴대폰

이라는 도구(Beale 2005)가 오로지 중독적 사용과 관련한 부정적인 심리 특성만을 가진다고 본다면, 무엇이 사람들로 하여금 휴대폰을 반복적으로 사용하게 만드는지, 관련된 다른 심리적인 특성은 없는지에 대한 종합적인 관점을 발달시킬 수 없을 것이다. 최근에는 사회적 네트워크 서비스(SNS)가 함께 발달하면서, 전술한 휴대폰 중독과 같은 부정적 측면 뿐 아니라, 이를 완화하고 조절하는 요인으로써 개인의 사회적 관계를 맺는 양식을 살펴보는 연구도 증가하고 있다. 사회지지 및 자아존중감이 휴대폰 중독을 조절한다는 연구(성준모 2014), 학교 및 가정에서 관계와 소통이 스마트폰 중독에 영향을 미친다는 연구(김희숙 외 2010), 가족 스트레스와 심리사회적 스트레스 대처능력이 상호작용하여 인터넷 중독을 조절한다는 연구(아영아 외 2010), SNS의 지지적 상호작용이 지각된 사회지지, 정서, 공동체 의식, 삶의 만족에 영향을 미친다는 연구(Oh 외 2014) 등을 고려했을 때, 휴대폰의 적절한 사용은 사회지지를 지각할 수 있는 통로가 되고, 사회교류와 정서지지도 높일 수 있다는 가설 설정이 가능하다.

이에 본 연구는 정신건강과 직접적으로 관련된 조현병 환자와 정상 성인을 대상으로 휴대폰 사용양상과 정신건강 및 지각된 사회지지와의 관련성을 알아보았다. 다만, 조현병과 같은 정신장애에서 인지적 쇠퇴를 경험하게 되고, 이에 어플리케이션의 사용이나 복잡한 조작 등에서 한계가 있다는 점을 감안하여, 휴대폰 사용 양상은 본래의 의사소통 기능을 대표하는 음성 통화와 문자 메시지로 제한하여 살펴보았다.

가설은 다음과 같다. 첫째, 조현병 환자와 정상성인 간에 정신건강과 지각된 사회지지, 휴대폰 사용량에서 유의한 차이를 보일 것이다. 둘째, 정신건강 수준과 사회지지와의 관계에서 휴대폰 사용의 조절효과가 있을 것이다. 마지막으로 휴대폰의 의사소통 기능은 사회지지에 영향을 미칠 것이다.

II. 연구방법

1. 연구대상

부산지역의 정신의료기관 및 사회복지시설을 이용하는 외래 조현병 환자 27명과 정상성인 27명을 대상으로 자기보고형 설문지를 실시하였다. 조현병 환자는 정신건강

의학과 전문의에 의해 진단 받은 지 12개월 이상 경과하고 1차례 이상 입원가료를 받았던 이들로, 신경학적 문제가 있는 환자는 배제시켰다. 참여자의 인구통계학적 변인을 보면, 조현병 환자(M=35.04, SD=10.03)와 정상 성인(M=29.88, SD=8.25)은 연령에서 차이가 있었으나, 교육연령(조현병 환자 M=12.92, SD=2.79, 정상 성인 M=14.38, SD=2.53)은 차이가 없었다. 참여자는 본 연구 목적과 검사에 대해 설명을 듣고 자필 서명한 동의서를 작성하였다.

2. 측정도구

휴대폰 사용 질문지. 음성/문자 사용량, 대상, 목적, 음성/문자 선호, 기타 기능 사용량, 입력된 전화번호 수, 월평균 요금 등과 관련된 정보를 질문지를 통해 얻었다.

사회지지. 사회지지를 측정하기 위해 애비 등(Abbey et al. 1985)이 개발하고 유성은(1997)이 변안한 사회지지와 갈등에 관한 척도를 사용하였다. ‘대상은 내가 하는 일을 고맙게 생각한다. 대상은 나를 존중해준다’ 등 주위 사람으로부터 얼마나 도움을 받고 있다고 느끼는지를 측정하는 8문항으로 구성되어 있다. 본 연구에서는 가족, 연인, 친구, 동료로 나누어 지각된 사회지지를 측정하였고, 신뢰도는 .85~.90이었다.

한국판 일반정신건강 질문지(Korea General Health Questionnaire-30; 이하 KGHQ-30). 정신건강 관련 심리문제를 측정하기 위해 신선인(2001)이 표준화한 KGHQ-30을 사용하였다. 일반인의 자료는 GHQ의 채점에 가장 일반적으로 쓰이는 ‘GHQ 채점법’을 적용하였다. 정신건강의학과 환자의 경우, 본래의 GHQ가 지난 2~3주 간에 변화한 심리적 상태를 그 이전 평상시와 비교하여 답하게 하는데, 현재의 정신건강 상태를 측정하기 위한 문항내용을 만성 환자들에게 적용할 경우에는 그 변화의 차이가 뚜렷이 나타날 수 없어, 대안적으로 굿차일드 등(Goodchild et al. 1985)이 제안한 CGHQ(Chronic GHQ) 방법을 적용하였다. 총 30문항이며 점수가 높을수록 정신건강이 양호하지 못함을 나타낸다. 본 연구에서 신뢰도는 .94였다.

3. 통계분석

정신건강, 지각된 사회지지, 휴대폰 사용양상에서 조현병 환자와 정상성인 집단 간에 차이가 있는지를 알아보기 위해 독립 t 검증을 사용하였다. 정신건강이 사회지지에 영향을 미치는데 휴대폰 사용이 그 수준을 조절하는지를 알아보기 위해 위계적 회귀분석을 실시하였다. 또한, 휴대폰의 의사소통 양상이 사회지지에 영향을 미치는지를 살펴보기 위해 다중회귀 및 위계적 회귀분석을 사용하였다. 자료 분석은 SPSS 16.0을 이용하였다.

III. 결 과

1. 조현병 환자와 정상성인의 정신건강과 지각된 사회지지, 휴대폰 사용 비교

조현병 환자와 정상성인 간에 정신건강 및 사회지지, 휴대폰 사용양상에서 차이가 있는지를 살펴보았다. 그 결과, 모든 변인에서 집단 간 차이가 있었다(표 1). 조현병 환자는 정상성인 비해 정신건강이 양호하지 못했고, 사회지지를 적게 지각하였으며, 음성통화의 수신/발신 횟수나 통화량, 문자메시지의 수신/발신 횟수도 유의하게 적었다.

〈표 1〉 일반건강질문지, 사회지지, 휴대폰 사용의 독립 t 검증

	집단	평균(표준편차)	t
KGHQ	조현병	18.19(5.07)	5.842***
	정상 성인	7.67(7.86)	
사회 지지	조현병	71.78(26.23)	-3.956***
	정상 성인	98.30(22.93)	

전화를 거는 수(건/하루)	조현병	1.93(1.57)	-3.149**
	정상 성인	7.33(8.78)	
전화를 받는 수(건/하루)	조현병	1.96(2.23)	-4.247***
	정상 성인	6.93(5.65)	
전화거는 양(분/하루)	조현병	7.56(8.17)	-2.449*
	정상 성인	29.26(45.33)	
문자메시지 보내는 수 (건/하루)	조현병	1.22(1.67)	-4.528***
	정상 성인	23.44(25.44)	
문자메시지 받는 수(건/하루)	조현병	1.63(1.67)	-4.918***
	정상 성인	24.33(23.93)	

* $\lt .05$, ** $\lt .01$, *** $\lt .001$

2. 정신건강과 지각된 사회지지와의 관계에서 휴대폰 사용의 조절효과

정신건강과 휴대폰 사용이 상호작용하여 사회지지에 영향을 미치는지를 살펴보기 위해 위계적 회귀분석을 실시하였다. 휴대폰 사용은 음성통화 횟수, 사용량, 문자메시지 사용 횟수 등을 z점수로 표준화하여 합산한 수치를 사용하였다. 정신건강이 양호하지 못할수록 사회적 지지를 적게 지각하였고(모형1), 조절변수인 휴대폰 사용이 증가할수록 사회적 지지는 증가했으며(모형2), 정신건강과 휴대폰 사용이 상호작용하여 이전 모형에 비해 설명력이 증가하고 있어(모형3), 조절변수가 상호작용 효과가 있음이 나타났다(표 2).

〈표 2〉 일반건강질문지와 휴대폰사용의 위계적 다중 회귀분석의 조절효과

		B	S.E. B	Beta	t
모형 1	(상수)	3.105	0.87		3.568
	GHQ	-0.24	0.057	-0.508	-4.248
		$R^2=.258$	F change=18.046***		
모형 2	(상수)	7.807	0.114		0
	GHQ	-0.535	0.133	-0.535	-4.018***
	휴대폰	0.017	0.033	0.068	0.51

		$R^2=.328$	$\Delta R^2=.070$	F change=12.452***	
모형 3	(constant)	0.173	0.124		1.399
	GHQ	-0.495	0.126	-0.495	-3.928***
	휴대폰	-0.051	0.04	-0.204	-1.283
	GHQ x 휴대폰	0.01	0.004	0.42	2.775**
			$R^2=.418$	$\Delta R^2=.090$	F change=7.701**

<.01, *<.001

3. 휴대폰 사용양상과 의사소통 기능의 선호도에 따른 사회지지 지각

먼저 휴대폰의 의사소통 기능(음성통화 및 문자메시지 사용)과 도구적 기능(게임, 음악청취 등)이 사회지지에 미치는 영향을 살펴보기 위해 다중회귀분석을 실시하였다. 그 결과, 두 변인의 투입으로 사회지지를 유의하게 설명하였지만, 의사소통 기능은 사회지지 지각에 유의한 영향을 미쳤고, 도구적 기능은 사회지지를 유의하게 설명하지 못했다(표 3).

〈표 3〉 휴대폰 사용의 다중 회귀분석

		B	S.E. B	Beta	t
모형 1	(상수)	0.103	0.154		0.668
	의사소통 기능	0.105	0.035	0.417	2.994**
	도구적 기능	-0.004	0.003	-0.195	-1.398
		$R^2=.154$	F change=4.546*		

**<.01

다음으로 휴대폰의 의사소통 기능 중 음성통화와 문자메시지를 선호하는 정도에 따라 사회지지의 지각 정도가 달라지는지를 살펴보기 위해 다중회귀분석을 실시하였다. 그 결과, 두 변인의 투입으로 사회지지를 유의하게 설명하는데, 음성통화 선호와

문자메시지 선호는 사회 지지의 지각에 유의한 영향을 주며, 종속변수에 대한 상대적인 기여도는 문자메시지 선호가 더 높았다. 이는 사회지지를 지각하는 데 음성통화보다 문자메시지 선호가 더 효과적이라는 것을 시사한다.

〈표4〉 기능선호의 위계적 다중 회귀 분석

		B	S.E. B	Beta	t
모형 1	(상수)	0.370	0.344		1.075
	음성통화 선호	-0.005	0.005	-0.154	-1.067
	$R^2=.024$		F change=1.139		
모형 2	(constant)	-2.519	0.849		-2.968
	음성통화 선호	0.022	0.009	0.618	2.499*
	문자메시지 선호	0.033	0.009	0.903	3.651**
		$R^2=.243$	$\Delta R^2=.219$	F change=13.330**	

* $<.05$, ** $<.01$, *** $<.001$

IV. 논 의

휴대폰은 의사소통적, 도구적, 오락적 기능 등 여러 가지 기능이 탑재되어 있지만, 무엇보다도 사람들 간의 의사소통을 연결해주는 중요한 역할을 한다. 따라서 휴대폰을 통한 의사소통이 많을수록 사회지지를 더 많이 지각할 수 있으리라는 가설 설정이 가능하다. 하지만 조현병과 같이 정신건강을 단적으로 나타내는 대상군에서 사회지지가 매우 중요함에도 불구하고, 전통적인 의사소통과는 다른 새로운 상호작용 방식에서 이를 살펴본 연구는 없었다. 이에 본 연구는 조현병 환자와 정상성인을 대상으로 휴대폰 사용이 정신건강과 사회지지를 지각하는 정도에 미치는 영향을 살펴보았다.

그 결과, 조현병 환자는 정상성인에 비해 정신건강이 양호하지 못했고, 지각된 사회지지의 수준이 낮았으며, 휴대폰 사용량도 적었다. 그리고 정신건강이 양호하지 못한 경우에도 휴대폰 사용의 조절효과를 통해 사회지지를 더 지각할 수 있었다. 즉, 휴대폰 사용은 사회지지를 많이 지각하여 정신사회 스트레스의 완충역할을 하는 것으로 보인다.

다. 그 중 휴대폰의 의사소통과 관련된 음성통화와 문자메시지 사용은 사회지지에 영향을 미쳤지만, 기타 기능의 영향은 유의하지 않았다.

한편, 휴대폰의 의사소통 기능 중, 대면 정보를 더 가지고 있는 음성통화는 사회지와 관련이 있을 것이라고 추정할 수 있다. 그러나 본 연구에서는 문자메시지에 대한 선호가 사회지와 관련이 더 많은 것으로 나타났다. 즉, 휴대폰의 의사소통 기능은 사회지에 영향을 주었지만, 도구적 기능은 영향을 주지 못했고, 의사소통 기능 중 통화보다 문자 선호가 사회지에 더 큰 영향을 미쳤다. 이는 모바일 상황에서 문자메시지를 사적인 일상대화과 사용자의 감정 상태를 표현하는 쪽지로 인식한다는 것을 의미한다. 또한, 음성메시지보다 문자메시지가 마음을 전달하는데 더 적합하며, 흔적이 남는다는 점에서 감동을 전달할 수 있다는 과거 연구결과를 지지한다(이수영 2003; 이종숙 2001; 이시형 외 2002).

그러나 정체성이 불완전하고, 관계를 맺고 있는 대상보다는 관계 자체에 의존적인 사람의 경우, 휴대폰의 집단 소비현상이 있다(박준범 2000). 특히, 휴대폰을 중독적으로 사용하는 사람들은 낮은 대인신뢰와 불안정한 정서를 보인다. 그래서 이들에게 상시 접촉 가능성의 특성을 지닌 휴대폰은 더 매력적으로 여겨진다(장혜진 외 2006). 또한 휴대폰 사용량이 많을수록 중독적 사용 가능성이 증가하고, 이에 따른 부작용도 현저하다. 본 연구처럼 휴대폰을 많이 사용할수록 사회지지를 풍부하게 지각할 수 있지만, 지나친 휴대폰 사용이 심리적 부적응을 야기할 수 있다는 선행 연구를 고려한다면, 적절한 휴대폰 사용 수준에 대한 추후 연구가 필요할 것으로 생각된다.

결론적으로 휴대폰 사용은 사회교류와 정서지지를 증가시키고, 심리사회 스트레스의 완충 역할을 했다. 따라서 휴대폰 의사소통 기능의 사용은 지각된 사회자원을 반영하는 지표가 될 수 있을 것이다. 그리고 휴대폰의 도구적 기능보다는 본래의 의사소통 기능, 특히 문자메시지 기능을 잘 활용하면 사회지지의 지각 정도도 증가할 것으로 생각된다.

본 연구의 제한점은 다음과 같다. 사례수가 많지 않아 일반화하기에 어려움이 있다. 또한, 결과에 영향을 줄 가능성이 있는 환자와 정상인 집단의 연령, 유병기간이나 직업, 교우관계 및 사회경제적 수준과 같은 변인들이 엄격히 통제되지 않았다. 그리고 사회지지를 예측하는 변인으로 정신건강과 휴대폰사용을 설정하였는데, 경험적으로

사회적 의사소통이 많은 사람이 휴대폰 사용량이 증가할 수 있어, 인과관계의 제한점이 있다. 임상적으로 휴대폰 사용과 같이 개인이 직접 통제할 수 있는 요인에 대한 개입이 중요하므로 조절효과를 알기 위해 본 연구에서는 인과관계를 설정하였으나, 이에 대한 추후 연구도 필요할 것으로 생각된다. 용어와 관련하여, 휴대폰 사용이 사회지지를 반영함을 나타낸다는 뜻에서 지표라는 단어를 사용하였으나, 측정과 관련되어 사회지지라는 개념을 평가하는 준거로써 기능하는지를 알아보지는 못한 제한점이 있다. 한편 최근 스마트폰과 SNS 어플리케이션이 폭발적으로 발달하고 있어, 본 연구에서 설정한 통화와 문자 메시지 이외의 기능들을 이용하면 정신건강이나 사회지지 효과도 달라질 가능성이 있다. 예를 들어 Ainsworth 외(2013)의 연구에서 정신증환자의 증상을 추적하기 위해 문자메시지와 어플리케이션을 사용했을 때, 앱을 사용하면 모니터링에 더 적은 시간이 소요되고 다양한 데이터를 모을 수 있으며 적극적인 피드백을 받을 수 있다고 나타났다. 따라서 스마트폰 앱을 이용한 추적관찰을 통해 적용 가능한 시사점을 찾을 필요가 있다.

참고문헌

- 과학기술정보통신부. 2015. 『2015년 인터넷중독 예방 및 해소 추진계획』.
- 과학기술정보통신부. 2018. 『2017년 12월 무선통신서비스 가입자 통계』.
- 김희숙·최연희·유성자. 2010. “고등학생의 자아정체성, 스트레스 및 인터넷 중독 정도와의 관계.” 『정신간호학회지』 19권 2호: 173-185.
- 나은영. 2001. “이동전화 채택에 영향을 미치는 이동전화 커뮤니케이션의 매체적 속성에 관한 연구 - 개인중심성, 즉시성 및 직접성을 중심으로.” 『한국언론학보』 45권 4호: 189-230.
- 나은영·박소라·김은미. 2007. “청소년의 인터넷 이용 유형별 미디어 이용 양식과 적응.” 『한국언론학회보』 51권 2호: 392-425.
- 박준범. 2000. 『청소년들의 이동전화 이용현상에서 나타나는 하위문화적 특성에 관한 연구』. 서울대학교 일반대학원 석사학위청구논문.
- 배영·박소라. 2005. “미디어 이용패턴과 사회적 관계: 어린이 집단을 중심으로.” 한국언론학회 가을철 정기학술대회, 265-266.
- 성준모. 2014. “대학생 스마트폰 중독에 영향을 미치는 스트레스 요인: 사회적 지지와 자아존중감의 조절효과.” 『정신보건과사회사업』 42권 3호: 5-32.
- 신선인. 2001. “한국판 일반정신건강척도(KGHQ)의 개발에 관한 연구 1 : KGHQ-20과 KGHQ-30을 중심으로.” 『한국사회복지학』 46권: 210-235.
- 아영아·정원철. 2010. “청소년의 학업 및 가족갈등 스트레스가 인터넷 중독에 미치는 영향: 스트레스 대처 능력의 조절효과를 중심으로.” 『청소년 복지연구』 12권 4호: 257-277.
- 엄태완. 2008. “빈곤층의 경제적 스트레스와 우울 관계에서 자기효능감과 사회적 지지의 효과.” 『정신보건과 사회사업』 28권: 36-66.
- 유성은. 1997. 『완벽주의적 성향, 사회적 지지, 스트레스에 대한 대처방식이 중년 여성의 우울에 미치는 영향』. 고려대학교 일반대학원 석사학위청구논문.
- 이수영. 2003. “이동전화 이용에 관한 연구: 음성통화서비스와 문자서비스간의 관계를 중심으로.” 『한국언론학보』 47권 5호: 87-114.

- 이시형·김학수·나은영·이세용·김진남·배진한·최병목. 2002. “청소년의 휴대전화 사용과 그 영향에 관한 연구.” 서울: 삼성생명공익재단 사회정신건강연구소.
- 이종숙. 2001. “치고 때리기, 문자서비스의 촉각성, 그리고 소통의 쾌락: N세대의 문자서비스 소비와 문화 읽기.” 한국언론학회 학술회의 논문집, 43-83.
- 이지영·유제민·현명호. 2008. “소방대원의 정서기능, 적극적 대처양식, 사회적 지지가 PTSD 증상에 미치는 영향.” 『한국심리학회지: 건강』 13권 2호: 373-389.
- 장혜진·채규만. 2006. “기술중독에 빠진 청소년들의 심리적 특성에 대한 연구 - 휴대폰 사용 중독을 중심으로.” 『한국심리학회지: 건강』 11권 4호: 839-852.
- 전현선·박정희. 2008. “지체부자유 학생이 지각한 사회적지지 및 자아개념 관계분석.” 『중북지체부자유교육』 51권 3호: 35-59.
- 정명숙·강석임. 2008. “정신장애인의 삶의 질에 영향을 미치는 요인: 자아존중감, 병식 및 사회적 지지를 중심으로.” 한국심리학회 연차 학술발표논문집, 562-563.
- 조윤경·백영민·김반야. 2014. “SNS 관계망에서의 상호작용은 어떻게 사회적 지지감을 형성시키는가? 관계망 유형에 따른 차이를 중심으로.” 『커뮤니케이션학 연구』 22권 2호: 5-31.
- 최일섭·고기숙. 2007. “정신장애인의 서비스이용, 가족지지, 사회적 지지가 삶의 질에 미치는 영향: 정신요양시설 거주 정신장애인을 중심으로.” 『정신보건과 사회사업』 27권: 331-359.
- 황경혜·유양숙·조옥희. 2012. “대학생의 스마트폰 중독사용 정도에 따른 상지통증, 불안, 우울 및 대인관계.” 『한국콘텐츠학회논문지』 12권 10호: 365-375.
- 홍은영. 2015. “대학생의 대인관계 문제와 우울의 관계에서 스마트폰 중독의 조절효과.” 『한국보건간호학회지』 29권 1호: 5-17.
- Abbey, A., Abramis, D. J., and Caplan, R. D. 1985. “Effect of different source of social support and social conflict on emotional wellbeing.” *Basic and applied social psychology* 6: 111-129.
- Ainsworth, J., Palmier-Claus, J.E., Machin, M., Barrowclough, C., Dunn, G., Rogers, A., Buchan, I., Barkus, E., Kapur, S., Wykes, T., Hopkins, R. S., & Lewis, S.. 2013. “A Comparison of Two Delivery Modalities of a Mobile

Phone-Based Assessment for Serious Mental Illness: Native Smartphone Application vs Text-Messaging Only Implementations.” *Journal of medical internet research* 15(4): 1-12.

American Psychological Association 1994. *Publication manual of the American Psychological Association* (4th ed.). Washington, D.C.: American Psychological Association.

Beale, R. 2005. “Supporting social interaction with smart phones. *Pervasive Computing, IEEE* 4(2): 35-41.

Cohen, S., and Hoberman, H. 1983. “Positive events and social supports as buffers of life change stress.” *Journal of Applied Social Psychology* 13: 99-125.

Oh, H. J., Ozkaya, E. and LaRose, R.. 2014. “How does online social networking enhance life satisfaction? The relationships among online supportive interaction, affect, perceived social support, sense of community, and life satisfaction.” *Computers in Human Behavior* 30: 69-78.

Thomé, S., Härenstam, A. and Hagberg, M.. 2011. “Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults - a prospective cohort study.” *BMC Public Health* <https://doi.org/10.1186/1471-2458-11-66>.

Can Use of Mobile Phone Be Applied as an Indicator of Perceived Social Supports in Patients with Schizophrenia?

Rinim Woo*·Seongjin Choi**

ABSTRACT

The purpose of this study was to investigate the effects of using mobile phone on mental health and perception of social supports. 27 patients with schizophrenia visiting community mental health and social rehabilitation facilities performed self-report surveys, and 27 normal adults went through the same procedure. The surveys included Questionnaire on Mobile Phone Use, Korean Version of General Mental Health Questionnaire-30(KGHQ-30), and The Scale of Social Support. Testing mean differences between groups and regression analysis were used for data analysis. The patients with schizophrenia showed worse mental health condition, lower perceived social supports, and less use of mobile phones compared with normal adults. Despite unfavorable mental health condition, the patients with schizophrenia could perceive more social supports through the moderating effects of mobile phone use. The communication function of mobile phone use had significant effects on social supports, but the instrumental function didn't. The preference for sending text message had better effects on social supports than the preference for phone conversation. The use of mobile phone could increase social interactions and emotional supports, and play a role as a moderator of psychosocial stress. The original function of mobile phone, communication, especially sending text message, would help increase higher level of perception of social supports than using mobile phone as an instrumental function.

Keywords: Mobile phone use, Social support, Mental health, Schizophrenia

투고일: 2019.01.12. 심사일: 2019.03.08. 게재확정일: 2019.03.14.

* Ph.D. Student, Department of Psychology, Gyeongsang National University.

** Professor, Department of Counseling Psychology, Tongmyong University.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.35>

【연구논문】

공공부문 프로그램평가를 위한 평가역량구축 방안의 모색

제 갈 돈*

논문요약

프로그램평가(program evaluation)를 위하여 새로운 지식을 습득하려는 학습에 몰두하거나 새로운 방법과 기법들을 발견하기 위하여 평가연구(evaluation research)라는 새로운 영역을 개척하려는 평가자들의 노력에 의하여 프로그램평가는 발전되어 왔다. 이런 노력의 결과로 프로그램평가는 학문 분야에서 자신만의 영역을 확보하게 되었고, 전문 직업으로서 정체성을 확립하게 되었다. 최근에는 프로그램평가가 정책의 결정, 기획 그리고 집행과 모니터링의 수준과 품질을 높이려는 목표를 성취하기 위하여 과학적 연구의 중요한 부분으로 진화하고 있다. 프로그램평가는 스스로 강력한 추진력을 가지고 발전하여 왔으며, 그 발전의 대가로 정책결정자, 프로그램 기획자 및 최고관리자들이 일정한 기준에 따라 자신들의 기능을 효율적으로 수행할 수 있게 되었다. 공공부문 프로그램의 평가역량을 구축하고자 하는 나라들이 공통적으로 접하게 되는 요소들이 존재하고 있고, 평가역량구축(evaluation capacity building) 노력의 성패는 그러한 요소들을 어떻게 관리하는가에 따라 결정된다. 본 연구는 그러한 공통적 요소들에 대해 경험으로부터 얻은 교훈과 이론적 논의를 전개하여 우리나라의 평가역량을 구축하기 위한 방안모색의 이론적 기초를 제공하고자 한다. 본 연구는 관심을 가진 독자들이 평가 실제로부터 평가역량구축 실재를 구분하고, 실제에서 평가역량구축이 무엇이며 상이한 조직 환경에서 어떻게 구축사업이 수행되는지를 이해하고, 그리고 평가역량구축의 언어, 개념모형 및 지표들을 구성하는데 도움을 주는 것이다. 이처럼 하드웨어와 소프트웨어를 아우르는 제반 평가시스템 구성 요소가 효과적이고 효율적으로 상호 작용할 때 그 조직 또는 국가의 평가역량이 우수하다고 할 수 있을 것이다.

주제어: 평가역량, 평가역량구축, 평가, 프로그램평가, 평가모형

* 안동대학교 행정학과, 교수.

I. 머리말

프로그램평가(program evaluation)는 오랜 발전 과정을 경험해 왔다. 일부 평가자들은 이미 증명된 믿을만한 방법을 선택하여 온 반면, 다른 평가자들은 새로운 지식을 습득하려는데 관심을 가지고 학습에 몰두하거나 또는 프로그램평가를 위한 새로운 방법과 기법들을 발견하기 위하여 평가연구(evaluation research)라는 아직 증명되지 않은 새로운 영역을 개척하려는 노력을 선택하여 왔다. 이런 노력들의 결과로 프로그램평가는 과학과 사회연구 분야에서 자신의 영역을 확보하게 되었고, 보다 중요한 사실은 프로그램평가가 독자적 분야 및 전문 직업으로서 자신만의 정체성을 확립하였다는 것이다. 최근에는 프로그램평가가 정책의 결정, 기획 그리고 집행과 모니터링의 수준과 품질을 높이려는 목표를 성취하기 위하여 과학적 연구의 중요한 부분으로 진화하고 있다. 프로그램평가는 스스로 강력한 추진력을 가지고 발전하여 왔으며, 그 발전의 대가로 정책결정자, 프로그램 기획자 및 최고관리자들이 일정한 기준에 따라 자신들의 기능을 효율적으로 수행할 수 있게 되었다.

프로그램평가는 이윤을 추구하는 조직들에게는 조직 운영의 능률성과 효과성이 성공적 관리의 측정을 위한 기준으로 재화와 서비스의 양과 시장성은 충분하지 않다고 알려져 왔다. 또한 프로그램평가 모형들이 매우 다양해지고 다양한 분야의 이익들을 위하여 활용되기 때문에, 전문 평가자들은 프로그램평가의 궁극적 목적은 제한되어서는 안 된다는 사실을 인식하고 있다. 한 가지 중요한 사실은 프로그램평가가 시장경제가 지배적이고, 시민사회 참여가 제도화되고, 이익집단들이 주도적이고, 그리고 비정부조직과 시민조직이 활발하게 성숙된 매우 분화되고 민주적인 사회에서 특히 보편화되어 왔지만, 아직도 조직의 전략과 장기 계획 속에 프로그램평가가 내재될 수 있는 방법으로 평가를 제도화하려는 욕구가 매우 강하다는 것이다. 이런 욕구는 프로그램평가를 조직의 기획에서 주된 관심 영역의 하나로 만들고, 따라서 기획에 상응하는 조직의 계획을 위한 적절한 평가역량구축(evaluation capacity building) 요소가 적소에 구성되어 있어야만 하는 것을 필요로 한다.

본 연구는 평가역량구축 프로그램의 주된 요소들에 대한 개관을 제공하기 위한 이론적 논의이다. 보다 구체적으로 본 연구는 궁극적으로 조직의 평가역량을 구축하는

데 무엇이 되고 무엇이 안 되는지를 밝힘으로써 미래의 조직 기획자들이 프로그램평가를 진실로 전략기획과 계획의 핵심 요소로 만들려는 욕구에 적절한 아이디어를 제공하여 프로그램평가 과정을 제도화하는데 필요한 기본적 체계를 제시할 것이다. 모든 나라는 각기 독특한 정치적·사회적 환경을 보유하고 있고, 따라서 그 환경에 적합한 방식으로 평가역량을 구축하기 위한 접근법을 개발해야 할 것이다. 그러나 많은 나라의 사례들을 비교분석해 보면, 평가역량을 구축하고자 하는 나라들이 공통적으로 접하게 되는 요소들이 존재함을 알 수 있다. 평가역량구축 노력의 성패는 그러한 요소들을 어떻게 관리하는가에 따라 결정된다. 본 연구는 그러한 공통적 요소들에 대해 경험으로부터 얻은 교훈과 이론적 논의를 전개하여 우리나라의 평가역량을 구축하기 위한 방안모색의 이론적 기초를 제공하고자 한다. 이를 위해 평가역량구축과 관련된 기존의 연구 결과들을 분석하여 공공부문의 평가역량구축에 관한 개념 정의와 의미, 평가역량의 구성요소와 상호관계 및 요소별 검토사항, 평가역량구축의 중요한 요소, 구축 방안 등이 논의될 것이다. 본 연구를 통해 얻어진 평가 및 평가역량구축과 관련한 교훈이 우리나라 공공부문 평가역량 구축을 위해 바람직한 정책방향을 제시할 것이다.

II. 평가역량에 대한 이론적 논의

1. 평가역량의 개념 및 의미

역사적으로 최소한 세 번에 걸친 큰 물결과 그 이외의 산발적 흐름을 통해 많은 국가들이 자국의 평가역량을 구축하기 위하여 노력해 왔다(Boyle et al. 1999). 제1의 물결은 1960년대로 미국, 캐나다, 스웨덴, 독일 등이 정부의 정책 및 프로그램의 효과를 제고하고 과정을 모니터링하기 위한 수단으로서 평가를 제도화하였다. 보건, 교육, 범죄, 주택, 복지 등 다양한 분야들의 정책 프로그램들을 기획하고 실행하는 정부 관료들과 프로그램 관리자들이 평가 결과의 주된 사용자들이었다. 제2의 물결은 1970년대 말부터 1980년대까지의 기간으로 노르웨이, 덴마크, 네덜란드, 영국, 핀란드, 프랑스 등이 평가제도를 적극적으로 도입하였다. 제도 도입의 주된 목적은 예산과

정을 통해 공공 책임성을 강화하고, 기존의 정책과 프로그램에 대한 정당화 논리를 재검토하며, 그리고 정부의 정치적 행위에 영향을 미치기 위한 도구로 사용하려는 것이었다. 평가는 주로 의회에 의해 주도되었고, 행정부에 대한 의회의 역할을 강화하는 수단으로 인식되었다. 제3의 물결은 1990년대 중·후반에 시작하여 지금도 계속되고 있는데, 유럽의 스위스와 아일랜드, 아시아의 한국과 인도네시아, 아프리카의 짐바브웨, 남미의 콜롬비아 등이 평가를 제도화하고 평가역량을 구축하는데 적극적이다. 세계은행은 평가역량개발팀(Evaluation Capacity Development Task Force)을 만들어 이들 국가들의 노력을 적극적으로 지원하고 있다.

평가역량구축에 관한 논의 중에 자주 사용되는 용어들의 정의와 유사 용어와의 구분을 분명히 하여 개념상의 혼란을 최소화할 필요가 있다. 이를 위해서는 평가의 개념, 정책·프로그램·프로그램의 관계, 산출물·효과·영향의 구분, 평가와 감사 그리고 평가와 진단의 차이, 평가와 모니터링의 관계, 메타평가, 평가제도, 평가역량 등의 내용을 간략하게 살펴볼 필요가 있다. 많은 학자들이 평가에 대한 정의를 내려왔다. Anderson(1984)은 평가를 체계적인 연구방법들을 적용하여 정책의 내용·집행·영향 등을 추정·사정·평정하는 작업으로, Dye(1976)는 특정한 정책의 과정과 결과를 이해하고 그 가치를 판정하는 사회적 과정으로, 그리고 Chelimsky(1985)는 프로그램의 설계, 실행, 그리고 효과성을 평가하기 위해 체계적인 연구방법을 적용하는 것이 평가라고 정의하였다.¹⁾ 최근의 논의는 프로그램평가가 구성에서 실행을 통한 파급효과에 이르기까지 프로그램 또는 정책 생명주기의 여러 단계들을 포괄할 수 있다는 것을 시사하고 있다. 따라서 평가는 정책·프로그램·업무 등의 수립, 집행, 산출, 결과 등을 평가기준에 따라 체계적으로 측정·평정하는 것이라고 정의할 수 있다(제갈돈·송건섭 2008; 이윤식 외 2006).

1) 또한 Wholey(1983)는 평가 그 자체보다 평가로 인한 개선 목적을 강조하여, 평가란 현재 집행 중인 프로그램이 주어진 목적을 효과적으로 달성할 수 있는가를 판정하는 것으로, 연구 설계의 논리를 사용하여 다른 상황적 요인들의 영향으로부터 프로그램 자체의 효과를 분리하여 현재 집행 중인 프로그램을 수정함으로써 프로그램을 개선하려는 의도적 노력으로 정의하였다. Rist(1990)는 평가의 방법론과 다룰 수 있는 문제의 종류 및 양 측면에서 내용과 범위가 뚜렷이 확장되어 왔음을 강조하며, 프로그램평가에 대한 주요한 가정은 그것이 엄밀한 방법론적 도구를 이용하여 프로그램의 결과 또는 효과를 평가하는 수단이라는 것이다.

감사는 약속된 규정이나 절차를 준수하고 있는지를 검증하는 절차이고, 평가는 규정 또는 절차의 준수 여부보다는 평가하려는 목적과 방향에 비추어 볼 때, 현재 평가대상이 어떠한 수준에 도달했는가에 대해 측정하는 것이다(신완선 외 2005). 감사란 검증하는 것이지 판단하는 것이 아니라는 말처럼 감사는 사실 여부의 검증에서 끝나는 반면, 평가는 측정 과정에서 또는 측정 후 결론을 도출하는 과정에서 어떠한 형식으로든 평가자의 판단이 개입될 수 있다는 점이 다르다고 할 수 있다. 평가와 감사와의 차이를 시행 목적과 방법론의 차이에 따라 구분할 수도 있다. 평가를 시행하는 목적은 효과성과 효율성의 진단, 개선 방향의 제시, 학습 등이라고 할 수 있는 반면, 감사를 시행하는 목적은 규정의 적법성 검증, 불법성 적발 시 경고나 교정 또는 처벌 등이라고 할 수 있다. 방법론에 있어서 평가는 피평가자의 협조가 필수적인 데 반해 감사는 강압적인 수사방식도 채택할 수가 있으며, 또한 평가는 사전에 설계된 평가 항목과 지침에 따라 시행되는 반면 감사는 감사과정에서 필요에 따라 감사 영역과 대상을 조정하는 경우가 많다(제갈돈·심광호 2010).

평가는 지속적인 모니터링으로부터 얻어진 성과정보의 주기적인 분석을 포함한다(Mayne and Zapico-Goñi 1997). 평가와 모니터링의 차이는 정도의 문제인데, 성과 모니터링은 결과보다는 일상의 업무에 초점을 맞춘다. 모니터링은 프로그램과 정책의 설계와 시행에 관심이 있는 반면, 평가는 프로그램이나 정책의 결과와 시행 필요성에 대한 근본적인 질문에 초점을 맞춘다. 그러나 실제에 있어서는 서로 겹치는 부분이 있고, 평가는 모니터링 자료에 크게 의지하곤 한다. 메타평가란 평가계획이나 진행 중인 평가 그리고 이미 완료된 평가를 다른 평가자가 평가하는 것을 의미한다(공병천 2008; 라휘문 2008). 상위 평가, 평가의 평가, 평가종합 등으로 불리기도 하는데, 메타평가는 여러 개의 기존의 평가에서 발견했던 사실들을 재분석하고 종합하여 기존 평가의 적정성을 평가함으로써 종합적인 교훈을 도출해 낸다(Lee 1999).

평가가 공식적으로 정부와 공공조직의 의사결정 과정의 일부가 되는 합법화 및 정당화의 과정이 제도적으로 명확할 때 평가가 제도화되었다고 할 수 있다. 평가 제도화는 정부의 공공정책에 대한 평가가 수행되는 규정과 절차, 조직체계의 구축 등 평가가 왜, 언제, 누구에 의해, 어떻게 이루어지는지를 규정하는 것이다. 평가 제도

화를 통해 평가가 불순한 특정 목적에 의해 이루어지는 것을 방지하고 최소한의 시행 빈도를 규정하게 된다. 본 연구에서는 다음과 같이 평가역량과 평가역량구축의 개념을 규정하여 논의에 적용하고자 한다. Boyle et al.(1999)의 논의에 의하면, 평가역량이란 기술, 지식, 경험 등의 인적 자원과 물질적·재무적 자원, 그리고 평가 실제의 정착 등을 의미하는 것으로 정의한다. 여기서 평가 실제란 평가의 정의, 평가연구의 설계, 평가 활동의 결과, 특정 정책에 영향 등을 의미하기 때문에, 평가역량구축이란 인적·물질·재무적 자원, 제도장치 등 평가역량의 구성요소들을 개발하고 강화시키는 모든 활동과 노력이라 할 수 있다(송건섭 2008). 따라서 평가역량을 계획에 따라 체계적으로 평가를 수행할 수 있게 하는 잠재적·실재적 힘으로 개인이나 조직이 효과적으로 평가를 수행할 수 있는 제반 능력이며, 이 능력을 개발하고 강화시키려는 노력이 평가역량구축이라 할 것이다(오철호 2010).

2. 평가역량의 구성요소

공공부문의 평가역량구축 방안을 모색하기 위해서는 우선 어떠한 요소들에 의해서 평가역량이 결정되는가를 검토해 보아야 한다. 그 요소들의 상대적인 중요성을 가늠해 보아야 하고, 요소 간 개선 노력에 있어 전후관계는 없는지도 살펴볼 필요가 있다. 중요하면서도 개선 노력에 선행되어야 할 요소들에 역량구축을 위한 초기 노력을 집중하고, 덜 중요하고 다른 요소들이 개선된 다음에 추가적인 개선 노력을 기울일 수 있거나 또는 장기적인 계획 하에 노력을 기울여도 평가역량의 심각한 불균형을 초래하지 않는 요소들에 대해서는 다음 시기의 개선 대상으로 미루어 두는 것이 효과적인 접근방법이 될 것이다.

공공부문의 평가역량을 결정하는 간접적인 환경적 요소들은 다음과 같다(Boyle et al. 1999). 첫째, 공식적인 통계, 경영공시 등 신뢰할 수 있는 자료들의 존재이다. 평가를 통해 유의미한 결론을 도출하기 위해서는 분석에 사용되는 자료가 객관적이고 신뢰할만한 자료여야만 한다. 만약 그러한 자료가 존재하지 않거나 수집될 수 없다면 이 부분에 대한 인프라 구축작업이 먼저 선행되어야 한다. 둘째, 사회과학적인 지식과 이해의 축적 정도가 문제이다. 평가 방법론과 실무는 정책과학, 경제학, 사회학, 경영

학, 행정학 등 사회과학 분야의 연구 성과에 의존하는 부분이 많다. 정부가 활용할 수 있는 사회과학 분야의 전문가가 국내적으로 충분하지 않다면 외국의 전문성을 활용하거나 사회과학 분야를 발전시키기 위한 노력이 필요하다. 셋째, 훈련된 분석 및 평가 인력의 존재 유무이다. 평가 방법론과 실무에 훈련된 최소한의 평가 전문가 집단은 존재하고 활용 가능해야 한다. 이들 전문가를 입법부 소속으로 활동하게 하거나 행정부 소속으로 활동하게 할 것인지 또는 정부 내부의 인력으로 운영하거나 외부의 인력으로 운영할 것인지는 차후의 문제이다. 역시 최소한의 전문가 집단도 존재하지 않다면 외국의 전문성을 빌려 내부 전문가들의 교육과 훈련을 선행해야 할 것이다. 넷째, 정치적·경제적 부정부패 존재의 정도이다. 평가가 의사결정의 질을 제고하고 정부 활동의 효과성에 대한 논쟁을 촉진하고자 하는 것이라면, 다양한 이해관계자들이 열린 마음으로 상호작용할 수 있는 환경이 필요하다. 부패했다는 것은 평가가 적절히 수행될 수 없음을 의미하고, 따라서 평가가 수행되더라도 평가 결과 자체가 왜곡되었거나 평가 수요자에 의해 무시될 것이다.²⁾

두 번째 요소는 조직으로 대부분의 나라들이 정도의 차이는 있지만 어떤 식으로든 정부업무를 평가하고 있다. 이는 정부업무의 평가를 담당하는 기관이 이미 설치되어 운용되고 있음을 의미한다. 그러나 평가관련 조직의 운용실태는 나라마다 상이하다. 행정부에 설치되어 있는 나라도 있고, 입법부에 설치되어 있는 나라도 있다. 조직, 인사, 재무 등 정부의 의사결정과 긴밀히 연계된 강력한 평가제도가 정착되어 있는 나라도 있고, 형식적이고 분절적인 평가에 머무르고 있는 나라도 있다. 중앙정부가 직접 평가를 시행하는 나라도 있고, 행정부처나 지방정부가 자율적으로 평가를 시행하는 나라도 있다. 이처럼 누가 평가의 목적과 내용을 기획하고, 누가 어떻게 실시하는지, 평가 결과를 어떻게 활용하는지, 그리고 평가담당 조직의 소속과 위상 및 운용실태는 역사와 상황에 따라 나라마다 상이하다.

평가관련 조직을 어디에 설치할 것인가 하는 점은 평가수요와 공급이 균형을 이루

2) 평가는 잘 조직된 거버넌스와 윤리경영을 필요로 하는데 이런 기반 하에서 만이 평가 결과에 대한 건설적인 분위기가 조성될 수 있기 때문이다(OECD, 1996). 반대로, 평가제도 자체가 건전한 거버넌스를 촉진하기 위한 장치이기도 한다. 그것은 평가제도가 정보 공개를 촉진하고 투명성을 제고시킴으로써 이해관계자들이 정부나 기관의 운영 상태를 감독하고 영향을 미칠 수 있는 환경을 조성할 수 있게 한다.

는데 중요한 역할을 한다. 민주적 과정으로서의 평가와 관리도구로서의 평가 사이에는 일정한 긴장이 있다. 민주적 과정으로서의 평가 기능을 강조하는 사람들은 평가 조직이 입법부 소속으로 설치되어야 한다고 주장한다. 이 주장은 평가의 효용성이 주된 관심사인데, 공공의 재원은 낭비 요소를 최소화하며 사용되어야 하고 이를 위해서는 견제와 감시를 통한 책임성 확보가 핵심이라는 것이다. 반면, 관리도구로서의 평가 기능을 강조하는 사람들은 평가 조직이 행정부 소속으로 설치되어야 한다고 주장한다. 평가를 통해 정부업무의 집행과정을 감독하고 정책과 정부 활동의 향상을 도모하는 것이 주된 관심사이기 때문이다. 이러한 차이점에 따라 평가체계의 유용성에 대한 평가 또한 달라질 것이다. 누가 옳고 그르고의 문제가 아니고, 문제는 현재 평가 제도를 구축하고자 하는 의도와 목적에 어느 방식이 가장 효과적인가 하는 점이다.³⁾

세 번째 요소는 제도로 평가가 수행되는 제도적 맥락을 고려하지 않고는 특정 국가나 정부의 평가 실재를 이해하거나 판단할 수 없다. 제도적 맥락은 평가가 특정한 목적을 위해서 실행되는지 아니면 체계적으로 수행되는지, 그리고 평가가 성과 위주의 공공관리에서 일정한 역할을 하는지, 한다면 어떻게 역할을 하는지 등을 결정할 것이다. 평가 결과가 자동적으로 의사결정 과정에 반영되는 것은 아니기 때문에 제도적 맥락이 중요하다. 평가를 수행하기 위해 구축된 평가 담당조직의 위상은 제도적 맥락의 구체적인 사항들에 의해 결정된다. 반대로, 평가 제도화의 성공 여부 역시 어느 정도는 평가를 담당하기 위해 설치된 조직의 위상에 의해 결정되기도 한다.

평가는 정책 기획, 인사, 조직, 예산, 감사 등 타 공공부문 기능과의 연계 속에서 활용될 때 그 가치가 더욱 제고된다. 정책이나 프로그램의 계획을 세우고 그 계획을 실행하기 위한 인사, 조직, 예산 등을 투입하면 그 결과 얻어지는 성과를 평가하고

3) 평가제도 구축의 의도와 목적은 정부의 기능에 대한 그 사회의 인식에 부합된 것이어야 한다는 것은 자명하다. 문제는 정부의 기능에 대한 인식이 시대에 따라 변할 수 있다는 점이다. 예를 들면, 1980년 이후 많은 나라들에서는 오랫동안 당연시 되어 왔던 중앙집권적이고 관료적인 정부는 더 이상 국민들이 요구하는 공공서비스를 제공하는데 적합하지 않다는 인식이 널리 확산되면서 민영화와 공공서비스의 상업적 접근을 요구하게 된 것이다. 이렇게 시대에 따른 국가 역할의 변화에도 불구하고 평가에 대한 요구에 지속적으로 부응할 수 있는 한 가지 방안은 입법부와 행정부 양자 모두에 평가 기능을 담당하는 조직을 설치하는 것일 것이다. 그러나 평가역량을 구축하는 초기에는 시대적 요구에 가장 적절한 평가체계와 조직적 위상을 신중히 고려하여 발전시키고 역사적인 진화과정을 밟는 것이 일반적인 접근법이다.

평가 결과에 따라 정책이나 프로그램의 중단, 축소, 지속, 확대 등 적절한 환류체계가 구축되어야 하는데, 평가는 이 환류를 위해 불가피한 기능이다. 환류가 새로운 의사결정으로 연결된다는 점에서 평가는 정부 의사결정의 중요한 요소를 구성하고 있다고 할 수 있다. 평가 결과의 환류체계, 즉 평가가 의사결정과 유기적으로 연계되어 있지 못하다는 것은 평가를 수행하는 주체와 그 결과를 필요로 하는 주체가 서로의 가치를 충분히 인식하지 못하거나 활용하지 못하고 있다는 것이 되고 결과적으로 평가 수요 자체가 약해지는 결과를 초래하게 된다. 평가가 연계되는 공공부문의 다른 기능 중 가장 중요한 하나는 정책기획이다. 정책기획이란 국정을 운영함에 있어 어떠한 정책, 프로그램, 프로그램 등에 얼마나 많은 재정, 인력, 시설 등을 투입하고, 어떠한 과정으로 추진해 나가야 할지의 우선순위를 결정하는 것이다. 대체로 이러한 정책의 기획기능은 예산기능과 긴밀히 연계되어 있다. Zapico-Goni(1993)는 변화와 불확실성의 시기에는 예산부처가 각 부처의 전략적 예산배정을 위한 역량구축을 선도해야 한다고 주장했다.

네 번째 요소는 평가의 공급과 수요로 공공부문의 평가역량은 평가공급과 수요가 동시에 향상될 때 발전할 수 있기 때문에 평가역량구축을 위해서는 평가공급과 수요를 동시에 개발하는 것이 효과적이다(노화준 2008). 평가관련 조직을 어디에 설치할 것인가 하는 점은 평가수요와 공급의 구성과 균형을 이루는데 중요한 역할을 한다. 평가공급과 수요에 관한 문제에 대하여 세계은행(World Bank)(1994)는 공공부문의 평가역량을 효과적으로 구축하기 위해서는 재무 및 정보인프라 구축의 필요성과 기회요인을 식별함과 더불어, 평가에 대한 수요와 공급에 영향을 미치는 요소들에 대한 점검으로부터 시작하여야 한다고 보고하였다. 만약 평가역량은 존재하지만 평가 결과에 대한 수요가 없다면, 평가기능은 미약한 상태로 존재할 것이다. 만약 수요는 존재하지만 평가역량이 부족하다면 평가는 적절하게 수행되지 못할 것이다. Picciotto(1995) 역시 평가공급과 수요의 관계와 관련한 구조적인 문제에 대하여, 공공기관의 관리자와 직원들이 평가 결과를 타당성 여부와는 무관하게 체계적으로 고려할 것이라는 보장은 없다고 하였다. 평가의 중요성에 대한 인식이 자동적으로 필요한 평가 조직체계의 구축으로 연결되지도 않는다. 평가가 그 본연의 잠재성을 잘 발휘하기 위해서는 독립성과 건설적 관계 유지를 통해 조직의 지속적인 학습과 업무 개선에 기여할 것이

요구된다. 이와 같이 평가수요와 공급의 연계는 평가 결과의 활용 측면에서 매우 중요하다. Bemelmans-Videc et al.(1992)이 지적하듯이, 평가 전문성의 보급과 실질적인 활용은 전문성의 공급과 수요가 상호작용할 때 비로소 나타난다.⁴⁾

평가공급은 평가를 실행하고 제도화하는데 요구되는 공급적인 측면에서 필요한 요소들로 평가방법론 및 지표체계, 평가에 필요한 경제적 자원, 평가자의 전문적 기량 등을 의미한다. Boyle et al.(1999)는 평가역량구축을 위한 인적 자원의 역량구축 필요성을 강조하면서, 평가자는 어떠한 자질을 갖추어야 하는지, 또 정부는 어떻게 평가자를 육성하여야 하는지 등을 논의하였다. 평가자에게 요구되는 기술이나 역량은 평가 환경의 변화에 따라 달라진다. 정보기술의 급격한 발전은 평가 활동에도 많은 변화를 초래할 것이다. 정보기술의 발전에 따른 변화는 정보기술 자체에 의한 변화와 다른 전문분야가 정보기술을 매개로 평가 영역으로 확산됨에 의한 변화로 대별된다 (Smith 2002). 정보기술 자체에 의한 변화는 모니터링을 위한 자료 수집과 평가가 온라인을 통해 실시간으로 이루어지고, 평가 결과를 제시하는 방법도 기존의 책자 제작에서 전자적 형태, 음성형태, 영상형태 등으로 다양화되며, 평가자가 평가 결과를 평가 수요자에게 전달하고 확산시키는 방법이 다양해지는 것 등을 포함한다. 다른 전문 분야의 침투에 따른 변화는 뉴스그룹 등을 통해 전 세계에 퍼져있는 평가 전문가들은 물론 다양한 분야의 전문가들이 평가 실체에 대한 토론과 정보 공유를 일상화하고 있고, 인접 학문의 지식과 경험이 빠르고 융합되고 있다는 점에 기인한다. 이는 이론적으로나 실무적으로 평가기반을 확대하고 심화시켜 줄 것이다.

평가수요를 제고시키는 방법은 다음과 같이 인센티브에 의한 방법, 벌칙에 의한 방법, 이를 혼용하는 방법, 문화를 조성하는 방법 등이 있다(Toulemonde 1999).

4) 그러나 Mayne(1994)는 개념적으로 단순한 이 관계가 실제로 있어서는 잘 인식되지 못하고 있다고 지적하면서, 상당한 평가 실체와 대부분의 평가 연구는 좋은 평가의 공급이 그 정보에 대한 수요와 결과의 활용을 창출한다는 공급 위주의 가정에 기반하고 있다고 한다. 반대로 수요 위주의 관점에서는 만약 평가정보에 대한 수요가 명확하다면 그 간격을 메울 수 있도록 평가가 이루어질 것이라고 주장한다. 공급 위주이든 수요 위주이든 어느 한쪽도 만족스러운 답을 주지는 않으며, 공급과 활용에 영향을 미치는 수요의 종류도 상이하다. 민주적 과정으로서의 평가 수요와 관리 도구로서의 평가 수요가 다를 것이며, 이 차이에 의해 평가 기능이 어느 조직에 위치해야 할지, 조직학습에 어떻게 영향을 미칠지, 누가 어떻게 결과를 활용할지 등이 결정될 것이다(Boyle et al., 1999).

인센티브에 의한 방법 중의 하나는 정책, 프로그램, 프로젝트 등의 예산편성 시 평가를 위한 예산을 별도로 책정하는 것이다. 공공기관의 특성상 평가를 위해 책정된 예산이 있으면 그 예산을 사용하고자 하는 유인이 생기게 되고, 따라서 평가용역 발주의 수요가 형성되는 것이다. 이때 주의할 점은 평가에 대한 수요는 다른 행정 및 관리상의 수요보다 약한 것이 일반적이므로 다른 용도와 함께 포괄적으로 책정된 예산은 평가수요로 연결되지 않을 가능성이 있다. 인센티브에 의한 방법 중 다른 하나는 평가 전담인력을 배치하는 것이다. 인력 배치는 예산배정과 유사한 측면이 있지만 다른 점은 평가 전담인력이 더욱 적극적으로 평가수요를 창출한다는 것이다. 일단 평가 전담으로 업무분장이 이루어지면 그 인력은 평가 전문성의 고양, 평가방법론의 개발 등 평가역량을 향상시켜 나가는 노력과 함께 기관장 혹은 평가 수요자의 요구가 무엇인가를 파악하고 그에 맞는 적절한 평가를 공급하기 위해 노력하게 된다.

Ⅲ. 공공부문 평가역량구축 방안

1. 평가역량구축의 중요한 요소

평가는 프로그램이 시작되고 진행되고 혹은 종료되어야 하는지의 여부를 결정하는데 그리고 만약 이미 집행 중에 있는 프로그램의 강도와 범위를 강화, 유지 또는 최소화하여야 하는지를 결정하는 성공적인 도구로 인정받아 왔다. 어떤 면에서 평가는 프로그램의 산출과 결과가 바람직한지 아닌지의 여부 그리고 어떤 과정들이 조직의 목적과 목표들의 성취에 기여하는지의 여부를 결정하는 수단으로 알려져 왔다. 아직도 일부 평가들은 공공 자금의 지출에서 법적·도덕적 책임을 조장하기 위한 감시자의 역할을 수행함으로써 공공부문에서 예산의 투명성을 확보하는 수단으로 운영되고 있다. 그러나 평가과정과 평가자들은 조직의 강점과 약점을 결정하는 분야에서 그리고 환경에서의 위협과 기회를 식별하는데 있어 성공적인 역할을 인정받아 왔지만, 지식 산출에서의 역할, 정책결정에의 영향 그리고 장기 및 전략기획에의 유용성에서는 아직 완전하게 활용되지 못하여 왔다. 그러므로 평가과정과 평가자들은 추가적으로 프로그

램평가에 관한 적절한 이해를 넓히고, 지속적으로 지식기반을 확충하고, 평가연구들을 통하여 효과적인 교육프로그램들을 촉진하며, 프로그램평가에 조직 프로그램과 정책들을 조정하고, 그리고 미래의 도전에 대처할 수 있도록 민간과 공공조직 모두의 평가역량 규모를 증가시키고 강화시킬 필요가 있다.

평가역량구축의 첫 번째 중요한 요소는 적절한 평가모형의 선정이다. ‘적절성’은 주관적인 개념이지만 우리들은 항상 기존의 일반적 조건들에 적용 가능한 기준들을 설정할 수 있고 동시에 효과성을 감소시키지 않고도 고유의 상황들에 반응할 수도 있다. 다음의 질문들이 적절한 평가모형의 선정이라는 목적 달성에 도움을 줄 수 있을 것이다. 먼저 평가모형이 다음과 같은 기본적 질문에 대답하도록 설계되어 있는가이다. 조직, 프로그램 혹은 정책을 개선할 여지가 있는가? 어떻게 할 수 있고 또 비용은 얼마인가? 조직에서 무엇이 일어나고 있는가? 프로그램 혹은 정책들이 이전 기존 조건들에 적절한가? 정책들이 실현되고, 프로그램들이 집행되고 그리고 목표들이 성취되고 있는가? 평가연구를 관리하는 최적의 시간은 언제인가? 평가모형이 효과적이기 위해서는 우선 언어가 단순하고 목적이 분명해야만 한다. ‘단순성’은 상이한 경험, 배경 및 다양한 견해와 훈련을 가진 광범위한 영역과 이해당사자들에 의해 쉽게 이해될 수 있다는 의미이다. 고도의 기술적인 평가 도구 혹은 보고서는 모든 필요한 정보를 수집해야 하는 평가의 목적을 손상할 것이며, 만약 사용되는 언어가 익숙하지 못하면 자신들과는 관련이 없다고 느낄 수 있는 지역사회의 지원을 얻기 어렵기 때문에 단순성에는 최고도의 주의가 고려되어야만 한다. 그 이유는 전문가 집단들이 언어를 이용하여 다른 사람들에게 권력을 행사할 수 있기 때문이다. 둘째, 적절한 평가모형은 시간제약적 연구이다. 이해당사자들이 수행되는 평가의 결과를 기다려 줄 수 있을까? 만약 그렇다면 정확하게 그들이 기대하는 것은 무엇인가? 그들에게 진행 중인 프로그램의 품질을 향상시킬 수 있는 대안들을 제시하는 형성평가(formative evaluation)를 제공하거나 혹은 기존 프로그램의 종결을 준비하는 총괄(summative) 연구를 제공할 수 있는가? 또는 연구가 단순히 수행된 과정들의 장점만을 강조하거나 프로그램의 결과에만 초점을 둘 것인가? 이해당사자들이 평가 후 무엇이 도출될 것인가에 대하여 적절하게 평가하고 중국적으로 결정하기 위하여 그들에게 주어진 시간은 별도로 한다. 셋째, 적절한 평가모형은 책임자를 식별한다. 평가역량을

구축하려고 하는 평가모형은 내부와 외부 평가자들 모두가 한 팀으로 일할 수 있는 기회를 제공해야만 한다. 만약 평가가 외부 전문가들을 제외시킬 것을 고려한다면, 조직이 구성원들의 평가 기술을 향상하고 평가를 프로그램의 일부로 포함할 이유가 없다. 그렇지만 관리자들은 조직과 외부 평가자들과의 관계를 공백 기간으로 취급하는 대신에 이런 기회는 프로그램의 좋은 장치로 그 속에서 내부 및 외부 평가자들은 연속체에서 공존할 수 있으며, 그럼으로써 다른 사람들을 경쟁자로서 잘못된 인식을 가지고 배척하려고 노력하기 보다는 바람직한 평가 결과를 제안하려는 목적의식을 가지고 다른 사람들의 문제를 인수하려고 노력할 것이다. 넷째, 적절한 평가모형은 평가하기 어려운 사회이슈들을 위한 방법을 제공할 것이다. 형성평가와 과정평가 기법들이 가진 한계에도 불구하고, 적절한 평가모형은 유용한 자원이 희소하다는 현실을 고려하여 실제 생활에서의 프로그램들을 평가할 수 있도록 실제적 장치와 접근법들을 결합하는 방법으로 두 가지 기법들에 약간의 조정을 시도하여 평가의 적실성을 성취함으로써 이 기법들을 활용할 수 있다. 다섯째, 평가 접근법이 없는 것이 오히려 완전하다는 사실이 지적되어야만 한다. 메타평가 접근법은 평가과정을 기초하고 있는 철학을 발견하려고 노력하는 과학적 의문들에 특별한 해결책을 보증하려는 접근법이다. Scriven(1972:1-4)은 평가자들은 프로그램 설계와 행정가들이 정의해 놓은 목적을 무시하고 발생할 것으로 기대되는 프로그램의 효과에 집중해야 한다고 제안했다. 여섯째, 적절한 평가모형은 다음의 조건을 만족해야 한다(Duignan 2002, 180-181): ① 평가의 신비성을 풀려고 시도함으로써 사회정책 영역의 모든 수준에서 평가가 이해되고 실행될 수 있다; ② 평가는 프로그램의 모든 주기에 걸쳐 시행되고 결과 평가에만 제한되지 않음을 강조하는 평가 용어들을 사용한다; ③ 평가는 내부 및 외부 평가자들 모두의 역할을 허용한다; ④ 평가는 평가하기 어렵고, 단지 이상형이 아닌 실제 세계의 프로그램들 그리고 대규모, 고비용 및 외부평가의 설계를 위한 방법들을 가진다; ⑤ 평가는 목표배제, 임파워먼트 등 평가의 메타접근법에만 특별한 관심을 주지 않는다; ⑥ 평가는 사회정책을 위하여 증거 기반의 측면에서 실제로 전달할 수 있는 것에 대한 이해에 기초한다.

평가역량구축의 두 번째 중요한 요소는 조직의 평가 프로그램들과 활동들을 조직화할 수 있는 기술의 향상, 기능적 시스템의 기반, 그리고 적절한 구조의 식별이다.

이상적으로 시스템, 구조 및 기술이 조직의 모든 수준에서 이용 가능하도록 만들어 지는 것이 가장 바람직하다. 그러므로 조직에 장착된 평가시스템이 가진 유의성을 모든 구성원들이 인지하는 것은 그들이 평가연구를 환류 메커니즘, 상황분석 도구, 예측장치, 변화 자극, 안정성 수단, 그리고 전략 및 장기 기획과 정책 형성을 위한 투입 자료로 적절하게 활용하는데 필수적이다. 성공적인 조직은 조직의 발전 수준만큼 복잡한 평가모형, 평가 작업에 몰입할 수 있는 멤버십, 그리고 조직 관리와 발전에 필수요소로 평가를 통해 정확한 질문을 하고 바람직한 대답을 제공할 수 있는 직원을 지속적으로 진화시킬 수 있어야 한다.

조직의 평가역량을 구축하는데 기여할 세 번째 요소는 의사결정자, 평가자 그리고 이해당사자들이 즉각적으로 행동할 가치가 있는 평가 이슈들을 식별하는데 도움을 줄 수 있는 정책과 기준들을 형성하는 것이다. 무엇이 긴급성과 즉시성을 구성하는지의 이슈는 설명되어야만 하는데, 왜냐하면 만약 평가 기획자들의 우선순위가 관리자의 우선순위와 경합하거나 아니면 보다 중요하게 만약 조직의 평가 직원의 주된 관심이 조직의 실제 평가 요구와 경합하게 되면 갈등이 발생할 것이기 때문이다. 이 같은 갈등은 평가자들, 기획자들 및 집행자들 사이에 대다수의 공감대는 없지만 정확하게 잘못 되는 것보다 대충 정확한 것이 좋다는 합의를 의미하는 것은 아니다. 최소의 방법은 평가 자원은 가능하면 가장 효과적이고 능률적인 방법으로 지출되어야 한다는 것이며, 가장 최선의 방법은 만약 필요하지 않다면 모두 다 소비하지 말고 피할 수 없을 때는 현명하게 소비하는 것이다. 결국 작업의 우선순위를 원활하게 할 이 같은 기준들을 형성하는 것은 평가자들이 평가역량을 구축하는데 많은 기여를 하게 될 것이다.

2. 평가역량구축의 방안

조직에서 평가역량구축이 필수적이라면 구축에 필요한 평가인프라, 평가모형 개발, 기술 향상, 그리고 평가역량구축 프로그램에 관한 공통된 이슈와 관심을 조성하려는 시도를 서둘러야 하며, 그 시도는 과학적 탐구의 전통에 기초하여야 한다. Zucker(2004, 2-3)는 과학적으로 기초한 연구를 수행할 때 다음의 지침이 평가연구

의 수행에 많은 기여를 할 것이라고 확신하였다. ① 관찰과 실험에 기초한 체계적·경험적 방법들을 사용한다. ② 진술된 가설들을 검증하고 일반화된 결론을 정당화하는데 적합한 엄격한 데이터 분석을 포함한다. ③ 평가자들과 관찰자들 사이, 다중 측정과 관찰 사이, 그리고 동일한 혹은 상이한 조사자들의 연구 사이 모두에 적용되는 신뢰적이고 타당한 데이터를 제공할 수 있는 측정 혹은 관찰 방법들에 의존한다. ④ 개인, 집단, 프로그램 혹은 활동들이 상이한 조건들에 할당되는 무작위할당실험이 바람직하지만 그것이 어려울 상황에서는 내적조건 혹은 교차조건 통제를 가진 다른 설계의 실험 또는 준실험설계를 사용하여 평가한다. ⑤ 실험 연구들이 재생가능하고 적어도 연구의 발견들을 체계적으로 구축할 수 있는 기회를 제공할 수 있도록 충분히 상세하고 분명하게 제시되는 방법을 확보한다. ⑥ 비교적 엄격하고, 객관적이고 또 과학적인 검토를 통하여 전문 논문집이나 독립적인 전문가들에 의해 연구 결과가 수용된다. Forsyth(2004, 8)는 모든 과학적 탐구의 지침이 될 여섯 가지 연구원칙들을 제시하였다. ① 경험적으로 조사할 수 있는 유의한 질문들을 형성한다. ② 연구를 관련된 이론과 연계한다. ③ 질문에 대한 직접적 조사가 가능한 방법들을 사용한다. ④ 논리적이고 분명한 추리의 연결고리를 제공한다. ⑤ 연구들을 재생하고 일반화한다. ⑥ 전문적 검증과 비판을 장려하기 위하여 연구를 개방한다. 더욱이 조직의 지속적인 평가역량구축을 확보하기 위하여 조직의 임무-비전-목적-목표의 진술에 적절하고 유의한 정책들을 형성함으로써 평가역량구축이 제도화되어야만 하고 그 제도화에 기여할 정책, 절차 및 장치의 구성을 파악한다.

1) 평가인프라 구축

평가인프라는 상호 연결된 구조적 요소들의 집합으로 각 구성요소들은 서로 독립적으로 존재하고 기능하는 방법으로 전체 시스템과 운영을 위한 틀을 제공한다. 평가인프라는 다음과 같은 제안된 요소들로 구성된다. 첫째, 평가 팀은 숙련되고 몰입된 전문가들로 구성되고, 지속적인 훈련과정을 통해 창조적이고 혁신적인 모형과 실제에 구체적으로 전문화된 팀이다. 대체 평가 팀들은 어떤 형태이든 기존 멤버들의 이직을 대비하여 항상 훈련되어 있다. 둘째, 최고관리자, 중간관리자, 강사, 정책결정자 그리고 잠정적인 이해당사자 대표들도 평가 훈련에 참여할 기회를 가져야 한다. 셋째,

지속적인 교육과 훈련 과정의 구성을 통하여 조직에서 선발된 구성원들이 경쟁적인 학부, 대학원 그리고 전문적 평가 경험과 학위과정에 참여하고 등록하는 것을 장려한다. 넷째, 평가기술의 재조정을 위하여 훈련을 필요로 하는 평가 연구자와 관심을 가진 직원들에게 기회를 제공한다. 다섯째, 전문적인 연구와 평가조직 간의 연계 및 파트너십의 구축이 필요하다. 여섯째, 조직의 관리정보시스템에 평가와 평가연구에 관한 문헌을 포함시킨다.

조직이 끊임없이 진화과정을 추구하는 것처럼 조직에서 사용되는 평가과정과 도구에 대해서도 조직은 평가 및 혁신에 필요한 자원과 욕구를 충족할 수 있도록 중요성도 인식하고 관심을 가져야만 한다. 이를 위해서는 다음과 같은 제안이 필요하다.

- ① 다양한 학문 분야의 연구자들 사이의 합의를 도출하기 위하여 학회를 통하여 다양한 분야를 포괄하는 평가연구에 메타분석 도입
- ② 다양한 형태의 평가연구 노력에 적용 가능한 다중방법 평가모형의 개선
- ③ 측정의 신뢰성을 확보하고 나아가 인과관계 증명에 관계자들의 자신감을 강화시킬 수 있는 내적 및 외적 타당도 위협요소들을 제거할 수 있는 기법의 개발
- ④ 학제적 평가교육 과정에 적합한 창조적 평가기법을 개발하기 위하여 설계된 개념 틀과 관점의 세심한 조정
- ⑤ 신뢰할 수 있고 바람직한 결과에 도달하기 위하여 필요한 시간과 노력을 줄일 수 있는 능률적이고 비용효과적인 접근법의 개발
- ⑥ 이해당사자, 평가 연구공동체, 관심을 가진 이익집단, 적절한 정책결정자 등 다양한 청중들에게 평가 방법, 발견 및 결과를 전파할 수 있는 창조적인 접근법들의 고안

2) 평가역량구축 프로그램

평가역량구축 프로그램 계획은 만약 능률성, 효과성, 활용된 결과, 조직의 운영과 기능적 능력 등에서 바람직한 변화가 실행되었다면 성공적으로 프로그램의 목적을 달성한 것으로 볼 수 있다. 평가역량구축 프로그램 자체에서 더 바람직한 것은 평가역량구축 프로그램의 능률성 수준의 정도에 따라 측정 가능한 지표들을 제공하는 것이

다. 다시 말하면, 이 지표들을 통하여 어떤 수준에서는 이 측정치가 측정되고 있는 조직의 프로그램과 활동들이 아니고, 프로그램의 목적 달성 여부를 결정할 수 없을 것이다. 마지막으로, 만약 평가 결과가 비교적 엄격하고, 객관적이고 그리고 과학적인 검토를 통하여 심사자들의 결절에 의하여 논문집에 게재되거나 혹은 독립적 전문가 패널에 의해 승인이 되어 왔다면, 우리는 평가과정은 조직의 평가역량구축에 기여하여 왔다고 인정할 수 있을 것이다(Zucker 2004, 3).

평가역량구축 프로그램의 성공적 실행을 제안은 다음과 같다. 첫째, 방법론에서 제안된 평가 설계는 양질의 지표들을 식별하고 제시하여야 한다(Thompson et al. 2005, 185). 모든 변수들에 대한 측정 점수의 신뢰계수들은 명백하고 합리적인 타당성을 가지고 사전 연구와 검증 매뉴얼에 기초하여 보고되어야 한다. 또한 이 계수들은 특정 연구를 위하여 수집된 데이터에 기초하여야 하며, 점수의 타당도는 연구 내에서 산출된 데이터에 기초하여 경험적으로 평가 되어야 한다. 둘째, 기술에서 제안들은 보다 효과적이고 능률적인 시스템의 구축을 위하여 학습을 강화하고 창의성을 장려할 수 있는 광범위한 범위의 기술을 어떻게 이해시킬 것인지를 제시하여야 한다. 또한 보다 새롭고 좋은 기술의 지속적인 출현 때문에 학습과정이 중지할 수 없는 것처럼 평가분야가 아직도 얼마나 방대한가를 적절하게 알리는데 도움을 줄 수 있도록 평가연구과정이 설계되어야 한다. 그렇지만 연구과정의 능률성과 효과성은 기술적 진보보다 우선한다는 사실이 지적되어야만 한다. 셋째, 기술이전에서 평가 결과에 대한 소통은 재고로 취급되어져서는 안 된다. 기획에서 소통은 평가가 계획되고 있을 때 시작되어야만 하며, 만약 계획이 시작단계에서 분명하고 상세하게 설명되면 소통이 잘 된 것으로 보인다(Posavac 1997, 247). 제안들은 평가 결과들이 어떻게 학자들의 논쟁, 학문 및 전문적 성장, 그리고 생산적 사회활동에 기여하는지를 제시하여야만 한다. 기술이전은 연구, 문헌검토, 파트너십과 네트워크 개발을 위한 훈련과 워크숍, 그리고 청중들을 위한 적절한 상호 학습경험을 허용하는 다른 형태들의 집행과 적용을 가능하게 할 수 있는 선도 영역의 수용에서 실현될 것이다. 넷째, 조직역량개발에서 평가연구계획은 조직과 협력기관의 평가역량구축에 기여한다는 측면에서 협력과 기대되는 결과의 범위와 본질에 따라 수혜자, 파트너 혹은 양자 모두로 구분되어야만 한다. 다섯째, 평가역량개발의 인간 요인에서 제안들은 어떻게 연구 활동이 지식 창출

을 촉진할 것이며 동시에 창조, 미래 평가자들을 위한 평가기술과 적성을 위한 지도 및 훈련, 그리고 새롭고 좋은 기술의 학습과 발견을 촉진하는가를 고려하여야만 한다. 결과적으로 계획된 활동은 평가과정에서 대표되지 못한 부문에 기회를 제공해야만 하며, 그럼으로써 연구자들은 인적 자원개발 목적을 위한 비학문적 기관들을 보다 전문적으로 확장할 수 있다.

지속적으로 평가역량구축 프로그램을 유지하기 위해서는 평가역량구축정책들이 다음과 같은 요소들을 포함하여 형성되어야만 한다(Duignan 2001): ① 평가인력 개발; ② 평가를 위한 우선 영역을 결정하기 위한 기준들; ③ 파트너들을 선발하는 기준과의 연계 및 컨소시움의 구축; ④ 연구에서 도출되는 선도 프로그램의 집행을 위한 영역 선발의 절차; ⑤ 평가자들과 고객들에 대한 윤리적 기준들; ⑥ 평가역량구축의 창도를 위한 소통.

IV. 맺음말

평가역량구축 과정과 실체는 프로그램평가 과정과 실체가 유사성을 보여주는 것과는 다르고 이들 사이의 구분은 지나치게 강조될 수 있다. 이 구분은 분명해질 필요가 있는데 이를 위하여 평가역량구축이 평가공동체 내에서 정당한 것으로 보이고 수용될 수 있는 가능성을 가질 수 있도록 분석과 논의를 개방해야 한다. 이 구분은 조직, 프로그램, 정부 그리고 공동체들이 평가와 그의 사용을 일상적인 실체로 인지하고 유지하는지 여부의 판단과 유지 방법을 위한 실질적, 즉각적 및 실제적 결과를 유도한다. 이 같은 구분에서 평가역량구축의 정의는 양질의 평가와 일상적인 평가의 활용을 실현하기 위하여 조직의 전반적 과정들을 지속적으로 창조하고 유지하려는 의도적 작업이라고 할 수 있다. 이 정의는 모든 프로그램 평가자들이 평가역량구축이 분명한 책임, 작업, 기술 및 행태를 가지고 정당한 전문적 역할로 보이기를 원하는 것만으로 평가역량구축 실체를 기대하지는 않는다.

프로그램평가 실무자들과 평가역량구축 실무자들 사이의 분명한 구분은 다음과 같이 설명할 수 있다. 민간 평가 실무자가 중간관리자들에게 평가기술을 가르치기

위하여 고용되었다면, 그 고용의 목적은 중간관리자들이 평가프로그램, 이해당사자 및 평가질문들을 식별하고, 평가질문에 대답하기 위한 가장 적절한 방법들을 결정하고, 데이터 수집도구들을 개발하고, 정보를 수집하고, 정보를 분석하고, 그리고 의사결정을 위하여 정보를 사용하는 등 기본적인 평가기술을 배우는 것을 돕기 위한 것이다. 훈련기간 동안 평가 관계자들은 다수의 프로그램을 위한 평가계획을 확인하고 개발한다. 훈련의 초점은 평가기술을 가르치고 중간관리자들이 보유하고 있는 기존 평가도구들에 새로운 평가기술을 추가한다. 이 훈련은 평가역량구축의 전형적인 방법으로, 여기서 평가 실무자는 중간관리자이며 어느 누구도 평가 작업이 유지되고 평가가 조직의 실체가 되도록 보장하는데 일차적 책임이 없기 때문에 엄밀한 의미에서 평가역량구축이라고 보기 어렵다.

다른 경우에서는 외부 평가자가 평가 실무자 역할과 평가역량구축 실무자 역할을 담당할 수 있다. 고객들은 프로그램의 형성적 및 총괄적 평가를 원한다. 평가 실무자들은 다른 이해당사자들을 식별하고, 형성적 및 총괄적 평가질문들을 확인하고, 그 질문들에 최선의 답변을 할 수 있는 방법들을 찾아내고, 그리고 이 프로그램 수행을 위한 평가계획을 개발한다. 평가 실무자들은 직원들과 함께 질적 및 계량적 데이터를 수집하고 즉시 환류함으로써 프로그램의 강점들을 확장하고 약점들을 파악할 수 있을 것이다. 여기서 평가역량구축 실무자들의 초점은 상황을 변경하고 유지하기 위하여 계획된 과정과 실제들의 의도적 행동시스템에 있으며, 그 시스템 속에서 양질의 평가와 적절한 활용이 하나 혹은 그 이상의 조직들과 프로그램들 내 또는 사이에서 통상적이고 진행 중에 있다. 다양한 평가의 새로운 계획들을 어떻게 통합하는지에 대한 의문이 평가 실무자들에게 제기되었을 때 평가역량구축은 프로그램평가보다 고객에 초점을 둔다. 고객은 지속적으로 품질향상위원회, 고객중심위원회, 전략기획위원회 등을 운영하고 있다. 외부 자금지원 때문에 프로그램 직원은 프로그램의 결과를 확인하여 왔으며, 평가를 통하여 무엇이 습득되었고 어떻게 실체를 개선하는데 사용되어 왔는가에 관한 정보를 포함한 평가 결과 데이터를 외부 자금 지원자에게 제공하여 왔다. 행정 팀은 형성적 및 총괄적 평가, 결과 평가, 지속적 품질개선 노력과 고객 초점 및 전략기획 노력을 완성하는데 필요한 평가 실제로부터 어떻게 배울 수 있으며, 또 이질적이고 관련이 없는 활동들을 수행하는 대신에 기관 스스로 학습조직이 될

수 있도록 이런 노력들을 어떻게 통합할 수 있는가를 알기를 원한다. 여기서 평가 실무가는 기관의 행정 팀을 위한 평가역량구축의 조력자가 되며 행정 팀은 평가역량구축 실무가가 된다.

조직의 관점에서 프로그램평가 실무가와 평가역량구축 실무가 사이의 구분이 확인될 수 있다. 프로그램평가 실무가 관점에서 조직 내 평가부서의 창설과 유지는 원칙적으로 사용 가능하고 한정된 작업에 초점을 둔다. 전형적으로 이런 관점은 다수 고객들이 사용하는 데이터 수집 도구를 표준화하고 평가 결과의 활용을 보장할 수 있는 연구 수행에 대한 고객들의 요구에 대한 반응을 의미한다. 이런 방법에 의한 평가부서의 창설과 유지는 일시적인 평가의 요구에 대한 반응으로 보일 수 있다. 조직은 실제적인 평가를 위하여 체계적으로 부서를 창설하는 보다 큰 목표를 소홀히 하고 평가부서가 조직 전반에서 정례적인 관습으로 운영될 수 있다. 그러나 평가역량구축 관점에서 초점이 평가 서비스를 위한 고객의 요구에 대한 반응에 있으며, 한편으로는 오늘의 업무가 어떻게 장기적으로 부서의 유지에 기여할 것인가에 있다. 이런 평가의 이중역할은 주어진 상황이 프로그램평가 실무가와 평가역량구축 실무가가 이중 초점을 가지고 각기 다르게 평가하는 것으로 생각하게 한다. 고객과의 상호작용에서 평가 실무가는 다음과 같은 질문을 하는데, 즉 어떻게 평가가 보다 좁은 연구를 유도하고, 보다 많이 사용되며, 또 다른 양질의 연구를 가능하게 하는데 기여하는가? 평가역량구축 실무가는 어떻게 특정 평가가 후속 연구들을 유도하고 활용되도록 조직의 학습과 발전에 기여하는가? 여기서 평가역량구축 실무가는 어떻게 개별 연구가 조직의 발전과 조직의 목표와 임무 달성에 연계되어 있는지를 생각하여야 한다. 여기서 제시된 세 가지 경우는 어느 것이 조직의 일상 업무에 가장 적합한지를 파악하고 조직 고유의 실제에 적용하도록 유도할 수 있다.

본 연구는 관심을 가진 독자들이 평가 실제로부터 평가역량구축 실재를 구분하고, 실제에서 평가역량구축이 무엇이며, 상이한 조직 환경에서 어떻게 구축사업이 수행되는지를 이해하고, 그리고 평가역량구축의 언어, 개념모형 및 지표들을 구성하는데 도움을 주는 것이다. 여기서 명심해야 할 사항은 평가역량구축은 평가역량구축 과정들을 촉진하는 아무런 책임이 없는 대다수 내부 및 외부 평가자들의 일상 업무 세계와는 구분된다. 대다수 평가역량구축 과정들은 사회 및 경제개발 활동들의 일부이고 정부

기관들을 포함한 공식 조직들에서 추진된다. 평가 실무자들은 여기서 일상의 운영을 이해할 것이지만, 그러나 쉽게 평가역량구축을 구현할 수 있는 조직의 개념 틀을 보유하고 있지는 않을 것이다. 그 개념 틀의 목표는 조직학습과 조직발전에 관한 개념화와 이론적 이슈와의 연계가 될 것이다. 평가역량구축은 쉽게 구축의 과정, 실제, 직업 오리엔테이션 그리고 실무자 역할에서 신비성이 제거된다. 평가자와 평가역량구축 실무자들 모두는 자료의 접근성, 흥미, 정보 등을 발견하여야만 한다. 보다 성숙한 평가역량구축 업무의 개념화와 품질은 세심하고 분명한 사정과 평가를 통하여 더욱 강화될 필요가 있다. 정책, 프로그램, 업무 등을 적절히 평가하여 국가 재정의 불필요한 낭비를 억제하고, 대국민서비스 향상을 위한 정부혁신을 지속적으로 이루어 나가기 위해서는 제도, 절차, 조직 체계 등의 정비와 같은 하드웨어적 노력과 그 제도를 운영하는 평가 인력의 전문성, 평가지표체계의 적실성 등과 같은 소프트웨어적 노력이 병행되어야 한다. 이처럼 하드웨어와 소프트웨어를 아우르는 제반 평가시스템 구성요소가 효과적이고 효율적으로 상호작용할 때 그 조직 또는 국가의 평가역량이 우수하다고 할 수 있을 것이다.

참고문헌

- 공병천. 2008. “상위평가에 관한 비판적 논고: 중앙행정기관 평가를 중심으로.” 『한국행정연구』 제14권 1호: 66-95.
- 노화준. 2008. “정부조직의 평가역량 구축방안.” 『행정논총』 제44권 4호: 373-403.
- 라휘문. 2008. “지방행정 혁신평가결과의 수용도에 대한 영향요인분석.” 『지방행정연구』 제21권 2호: 199-224.
- 송건섭. 2008. “정부업무평가제도하의 자체평가역량에 관한 실증분석.” 『한국정책과학학회보』 제11권 1호: 57-79.
- 신완선 외. 2005. 『평가의 기술』. 서울: 해바라기.
- 오철호. 2010. “논의의 초점: 평가활용, 평가역량 그리고 정책학습.” 『정책분석평가학회보』 제20권 2호: 1-12.
- 이윤식·김주환·김홍률·윤기석·심광호·박재신. 2006. 『정부성과관리와 평가제도: 주요 선진국 사례를 중심으로』. 서울: 대영문화사.
- 제갈돈·송건섭. 2008. “정부업무평가제도의 발전을 위한 메타평가모형 구축방안.” 『한국정책과학학회보』 제17권 1호: 1-27.
- 제갈돈·심광호. 2010. “감사와 평가의 비교분석과 그 연계를 위한 규범적 및 제도적 방안 연구.” 『한국행정논집』 제22권 1호: 27-54.
- Anderson, J. E. 1984. *Public Policy Making (3rd ed.)*. New York: Halt, Rinehard & Winston.
- Bemelmans-Videc, M., Rist, R., and Verdung, E. (eds.). 1992. *Carrots, Sticks and Sermons: Policy Instruments and their Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Boyle, R., Lemaire, D. and Rist, R. 1999. “Introduction: Building Evaluation Capacity,” in R. Boyle and D. Lemaire(eds.), *Building Effective Evaluation Capacity: Lessons From Practice*. New Brunswick, NJ: Transaction Publishers.
- Chelmsky, E. 1985. *Program Evaluation: Patterns and Directions*. New York:

- The American Society for Public Administration.
- Duignan, Paul. 2001. "Mainstreaming Evaluation or Building Evaluation Capability?" Presented at "Evaluation 2001" Conference of The American Evaluation Association, St. Louis, Missouri. November 7-10.
- Duignan, Paul. 2002. "Building Social Policy Evaluation Capacity." *Social Policy Journal of New Zealand* 19(December): 179-194.
- Dye, Thomas. 1976. *Policy Analysis*. Alabama: University of Alabama University Press.
- Forsyth, P. B. 2004. "Perspective: Educational Administration Research. Are We Too Broken?" *School Leadership News* Fall(8): 2.
- Lee, L. 1999. "Evaluation Coverage," in R. Boyle and D. Lemaire(eds.), *Building Effective Evaluation Capacity: Lessons From Practice*. New Brunswick, NJ: Transaction Publishers.
- Mayne, J. 1994. "Utilizing Evaluation in Organizations: The Balancing Act," in F. Leeuw, R. Rist, and R. Sonnichsen (eds.). *Can Government Learn? Comparative Perspectives on Evaluation and Organizational Learning*. New Brunswick, NJ: Transaction Publishers.
- Mayne, J. and Zapico-Goñi, E. (eds.). 1997. *Monitoring Performance in the Public Sector: Future Directions from International Experience*. New Brunswick, NJ: Transaction Publishers.
- OECD. 1996. *Ethics in the Public Sector*. Paris: OECD.
- Picciotto, R. 1995. "Introduction: Evaluation and Development." *New Directions for Evaluation* 67 Fall.
- Posavac, Emil J., and Carey, Raymond G. 1997. *Program Evaluation: Methods and Case Studies (5th ed)*. NJ: Prentice-Hall Inc.
- Rist, R. (ed.). 1990. *Program Evaluation and the Management of Government: Patterns and Prospects Across Eight Nations*. New Brunswick, NJ: Transaction Publishers.

- Smith, M. 2002. "Evaluation: Preview of the Future." *American Journal of Evaluation* 22(3): 281-300.
- Thompson B., Karen E. Diamond, Robin McWilliam, Patricia Snyder, Scott Snyder. 2005. "Evaluating the Quality of Evidence From Correlational Research for Evidence-Based Practice." *Council for Exceptional Children* 71(2): 185.
- Toulemonde, J. 1999. "Incentives, Constraints and Culture-Building as Instruments for the Development of Evaluation Demand," in R. Boyle and D. Lemaire(eds.), *Building Effective Evaluation Capacity: Lessons From Practice*. New Brunswick, NJ: Transaction Publishers.
- Wholey, J. S. 1983. *Evaluation and Effective Public Management*. Boston, MA: Brown and Co.
- World Bank. 1994. *Report of the Evaluation Capacity Development Task Force*. Washington, D.C.: The World Bank.
- Zapico-Goñi, E. 1993. "Many Reforms, Little Learning: Budgeting, Auditing, and Evaluation in Spain," in Gray, A., Genkins, B., and Segsworth, B. (eds.) *Budgeting, Auditing and Evaluation, New Brunswick, NJ: Transaction Publishers*.
- Zucker, Sasha. 2004. *Scientifically Based Research: NCLB and Assessment*. Harcourt Assessment, Inc.

Evaluation Capacity Building for Program Evaluation in Public Sector

Don Jaegal

ABSTRACT

Program evaluation has been developed by the efforts of evaluators on learning to gain new knowledge for program evaluation or on creating the field of evaluation research to discover new methods and techniques. As a result of these efforts, the program evaluation has gained its own domain in academic fields and established its identity as a profession. In recent years, program evaluation has evolved to be an important part of scientific research to achieve the goals of making decisions, planning, and improving the level and quality of implementation and monitoring of policies. Program evaluation has evolved with strong momentum on its own, and in return for its development, policy makers, program planners, and top managers can effectively perform their functions according to a given standard. There are elements commonly encountered among countries seeking to build up the capacity to evaluate public sector programs, and the success or failure of implementing them depends on how they are managed. This study aims to provide the theoretical basis for finding ways to build up the nation's evaluation capacity by developing lessons and theoretical discussions drawn from experience on such common elements. This study identifies the actual implementation of the capacity building from the real world of the evaluation, identifies what is building the evaluation capacity in practice and how the implementation programs are carried out in different organizational settings. When these components of the overall evaluation system, including hardware and software, interact effectively and efficiently, the organization or country is likely to have a superior capacity to evaluate.

Keywords: Evaluation Capacity, Evaluation Capacity Building, Evaluation, Program Evaluation, Evaluation Model

투고일: 2019.02.11. 심사일: 2019.03.07. 게재확정일: 2019.03.12.

* Professor, Andong National University.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.61>

【연구논문】

장애인 거주시설 종사자의 인권기반 복지실천 경험에 관한 연구*

전 지 혜**

논문요약

본 연구는 장애인 거주시설 종사자들이 인권기반 복지실천의 과정 속에서 어떤 경험을 하고 있는지 현상학적 연구를 통해 살피고, 경험의 내용과 의미를 파악하고자 하였다. 29명의 연구참여자를 의도된 표집으로 선정하였고 총 3개의 포커스그룹을 구성하여 면담을 실시하였다. 수집된 자료는 Colaizzi의 현상학적 방법으로 분석되었으며 연구결과 3가지 상위범주와 6개의 주제가 도출되었다. 장애인 거주시설 종사자들은 실존적 가치의 혼란(종사자로서의 목소리 상실, 존재 가치의 상실), 철학적 가치의 혼란(자립생활 철학에 대한 의문, 자기결정권의 본질에 대한 의문), 실천적 방향상의 혼란(자기방어적 실천, 기존의 실천방식에 대한 고민)을 경험하고 있는 것으로 나타났다. 즉 거주시설이라는 제약된 공간에서 자기결정권 존중으로 대표되는 인권적 복지실천을 해야 하는 복지패러다임의 전환이 종사자들에게는 혼란스러울 수밖에 없고, 이러한 상황은 패러다임 전환의 과도기 속에서 거주시설 종사자들이 겪어내야 할 과업으로 해석되었다. 이러한 논의를 바탕으로 거주시설 종사자들이 인권적 복지실천을 마련할 수 있는 정책적 대안을 제시하였다.

주제어: 인권기반 복지실천, 장애인 거주시설, 사회복지실천, 딜레마, 장애패러다임의 전환

I. 서론

한국의 장애인 복지정책은 2000년대에 접어들면서 장애인의 자립 생활을 중심으로

* 이 논문은 인천대학교 2017년도 자체연구비 지원에 의하여 연구되었음.

** 인천대학교 사회복지학과, 조교수.

재편되었다. 기존의 보호적 태도의 복지적 접근(welfare approach)에서 벗어나 장애인의 자립과 자기결정권을 존중하는 인권적 접근(human rights approach)으로 복지정책과 서비스의 기초가 변경되었다(김형식, 2008; Healy, 2008). 이러한 패러다임의 변화와 함께, 장애인에 대한 인권적 복지실천이 강조되면서 장애인 거주시설 내 장애인 인권문제는 사회적 이슈가 되었다. 2014년 신안군 염전노예 사건 및 인강원 사건이 보도되면서 장애인권에 대한 국민적 관심은 더욱 뜨거워졌고, 최근 1980년대의 형제복지원 사건이 재조명되면서 거주시설에서의 장애인에 대한 인권침해는 주요 기사로 다루어지고 있다.

시설 내 장애인에 대한 인권침해 문제에 사회적 관심이 모아지면서, 인권침해 실태 조사 및 연구들도 다수 수행되었다. 장애인 인권침해 실태에 관한 연구(남구현, 2005), 지역사회장애인의 인권침해 실태 연구(조한진 외, 2006), 장애인 거주시설의 인권연구(김미옥 외, 2008), 장애인생활시설에서의 인권침해 현황(임성택, 2012) 등이 대표적이다. 이 연구들은 2007년 장애인차별금지법의 도입 이후 국내의 장애인 인권문제가 시대적 화두임을 보여주고 있으며, 장애인이 인권이 보장되기 어려운 장애인 거주시설의 실태가 어떠한지 보여주었다. 결과적으로 거주시설에서의 인권적인 복지실천의 중요성이 강조되었으며, 거주시설의 내적인 변화를 위한 제도적 실천적 변화의 노력으로 이어졌다. 2011년 장애인복지법이 개정되어 신설 거주시설의 규모를 30인 이하로 하고, 30인 이상 시설에서 한 명의 장애인이 자립해 나갈 시에는 추가 인원을 받지 않도록 하였으며, 거주시설의 표준공간으로 가정과 같은 환경을 설정하였다. 이후 거주시설 일부는 거실, 화장실, 주방이 있는 형태로 거주 공간을 전환하였으며, 거주시설 종사자를 대상으로 한 인권교육도 의무적으로 실시되었고, 거주시설별로 인권지킴이단도 설치되었다.

현재 장애인 정책의 기초는 탈시설 자립지원 정책을 포함한 커뮤니티 케어라고 할 수 있다. 2019년 현재 보건복지부는 중증 장애인이나 노인이 지역에서 충분한 서비스를 받으며 지낼 수 있도록 지역사회 내 복합적인 돌봄체계를 갖출 것을 약속하고 시범사업을 추진하고 있다. 장애인 거주시설은 이러한 패러다임의 전환 속에서 탈시설 및 시설의 소규모화, 장애인의 자립지원 정책이라는 변화에 적응해 나가야 하는 것이다. 이는 전문가 중심의 재활패러다임에서 당사자 중심의 자립 패러다임

또는 인권기반 복지실천으로의 변화로 이해할 수 있으며 시설종사자들에게도 큰 변화라고 할 수 있다(김용득, 2016).

거주시설은 기본적으로 장애인에게 집이고 일상의 삶의 공간이다. 하루 24시간 동안 거주와 요양 서비스가 제공되기 때문에 거주시설 종사자와 이용자는 전체 일상을 공유하는데, 인권이라는 추상적 권리보장이 24시간의 생활의 장면에서는 종사자와 이용자의 의견 충돌의 모습으로 나타나기 쉽다. 장애인 거주시설 내 종사자들을 대상으로 한 최근의 연구들은 이러한 인권적 복지실천의 딜레마적 상황을 잘 논의하고 있다(김미옥·김경희, 2011; 김미옥 외, 2013; 박경수·장혜경, 2013; 조우연, 2013). 이 연구들은 ‘인권’이라는 추상적 개념이 일상의 삶 속에서 적용될 때에는 지극히 생활밀착적인 이슈가 되기 때문에(Ife, 2001), 인권기반 복지실천이 거대담론으로서 는 이상향이 될 수 있으나, 삶의 현실에서는 종사자와 이용자 사이의 힘겨루기 내지는 갈등을 야기하고 있다고 설명한다. 예를 들면, 잡곡밥이 싫다고 하는 장애인에게 영양 상태를 고려하여 잡곡밥을 강요할 수 있는지, 장애인의 임신 여부를 확인하기 위해 사용한 생리대 개수를 확인해야 하는지와 같은 지극히 사적인 인간으로서 보장되어야 하는 인권이라는 것이 생활에서는 매우 현실적인 이슈로 등장하고 있으며, 때로는 이것이 종사자와 이용자 사이의 갈등을 야기하기도 하는 것이다. 사회복지 실천가들이 장애인 이용자의 의견을 묻고 그에 따른 실천을 하는 것이 기본적 복지제공방식이라고 제시하고 있지만(정종화 외, 2009), 장애인의 욕구나 상황이 저마다 다르고 종사자의 가치관과 실천방식도 저마다 다르기에, 인권기반 복지실천이라는 방향적 지향은 현실 세계에서는 종사자와 이용자 사이의 갈등 내지는 종사자가 경험하는 딜레마로 나타날 수밖에 없다. 그럼에도 인권 담론과 관련된 논의에서는 복지실천의 이슈는 거의 다루어지지 않았고(조효제, 2009), 복지실천의 이슈를 다루는 데에서는 정책적 실천적 대응방안에 대한 논의가 이루어지지 못했다.

최근 몇 년간 이 주제를 다룬 연구들은 딜레마적 상황을 다루는 데 집중한 특성이 있다. 한 개 기관에서의 인권침해 사건 해결과정에 집중하거나(김경희, 김미옥, 2012), 거주시설 내 실천가만을 대상으로 일상에서의 인권 딜레마를 살피거나(박경수, 장혜경, 2013), 다차원 척도법을 활용하여 거주시설 종사자의 인권수행의 어려움의 개념적 종류를 살피는 것에만 집중하거나(김선주, 2016), 거주시설 내 사무국장의 인식에만

초점을 두는 등(김용진, 조원일, 2018), 각 연구마다 한계점이 있다.

복지 영역은 가치와, 정책과, 실천의 영역을 모두 아우르는 분야로서 복지의 가치철학과 정책적 기초의 변화가 현장의 실천 영역과 맞닿아 있다. 이에 장애인복지 영역에서의 인권기반 복지실천이라는 변화가 현장에서 어떻게 나타나고 있는지 알아볼 필요가 있다. 인권적 복지실천의 어려움이 단순히 패러다임의 전환에 따른 딜레마적 상황으로만 계속 나타나고 있는지, 자립 패러다임 또는 커뮤니티 케어가 강조되는 현시점에서 인권적 복지실천을 해나가는 종사자들의 경험은 어떠한지, 그 경험의 본질적 속성은 또한 무엇인지 알아볼 필요가 있다.

이에 본 연구는 인권적 복지실천이 강조되고 있는 장애인 거주시설 종사자를 대상으로, 장애인복지에서 종사자들이 인권기반 복지실천을 어떻게 경험하고 있는지 알아보고, 그 의미를 파악하고자 한다. 현상학적 연구방법을 적용하여 장애인거주시설 종사자의 인권적 복지실천의 경험 양상과 본질적 의미 파악에 집중하고자 한다. 인권적 복지실천은 우리사회의 장애인복지영역의 중요한 과제로 인식되고 있으므로, 이 연구를 통해 장애인복지 현장 속에서 종사자들의 경험을 심도있게 이해함으로써 향후 장애인복지의 인권적 실천을 위한 방안 마련에 기초자료를 제공할 수 있을 것이다.

II. 이론적 배경

1. 장애인복지 패러다임의 전환과 인권기반 장애인복지

장애인복지의 영역에서의 인권과 복지실천의 개념은 일반적인 사회복지실천에서 논하는 인권존중에 기반한 복지실천의 논리와 다소 다르게 전개되었다. 장애인 당사자 주의로 표현되는 장애인 당사자의 권리에 대한 보장, 특히 당사자에 의한, 당사자를 위한, 당사자의 자기결정권 복지정책 및 제공체계 전반에 대한 변화를 요구하는 장애인 인권운동은 기존의 복지제공방식이나 철학에 대한 저항담론으로 등장했다(이선우, 2009). 이들은 재활중심적 의료, 분리·배제적 특수교육, 시혜적 복지 등을 문제 삼으면서, 장애인의 비장애인과의 대등한 인권 보장을 위한 전체적인 장애인 정책의 구조

와 체계가 변화되어야 한다고 주장하였다. 이러한 요구는 국내적으로는 장애인차별금지법 입안과 국외적으로는 UN 장애인권리협약을 통해 관철되었고, 현재 장애인복지정책과 실천이 가야할 방향을 제시하고 있다. 인권중심적 장애인 정책은 지역사회 내 장애인의 자립생활과 사회에의 완전한 참여를 강조하였고, 장애인의 주거, 교육, 교통, 고용 정책 등의 영역에서 활동 지원 서비스에 이르기까지 실질적인 사회통합 정책들을 근간으로 하는 변화를 유도하고 있다. 장애인복지영역에서는 이러한 변화에 대한 요구를 수용하여 기존의 보호적 복지접근 방식을 벗어나 인권기반 복지로의 방향 전환을 시도하고 있으며, 간단히 정리하면 아래의 표와 같다.

〈표 1〉 장애정책 패러다임의 변화

구분	기존 복지 패러다임	새로운 복지 패러다임
명칭	보호적 장애인복지	인권기반 장애인복지
모델	개별적 모델	사회적 모델 강화를 통한 통합 모델 ¹⁾
목표	안정과 안전	기회의 평등 보장과 완전한 사회참여
정책수단	현금이전, 시설보호	법적수단 (장애인차별금지법 등 권익 보장)
실천방식	욕구사정 및 보호제공	당사자의 자기결정권 강조 및 권리보장

출처: 이소영(2011) OECD 국가의 장애정책패러다임의 전환, p.37 표를 수정함

장애정책의 패러다임의 변화는 장애인복지 패러다임이 변화하는 것을 의미하는데, 학자들마다 다르나 일관되게 권리기반 복지철학으로의 변화를 강조하였다. Doyle(1999)는 “자선기반에서 권리기반으로의 변화”라고 하였고, Waddington과 Diller(2002)는 “복지모델에서 민권모델로의 변화”라고 하였다. 새 패러다임은 장애

1) 장애에 대한 모델은 학자에 따라 여러 논의가 있으며, 기존 패러다임과 새로운 패러다임에 대한 이분법적 논의를 하기에는 무리가 있음. 단, 이해를 돕기 위해 장애정책 패러다임의 변화를 표로 제시하였으며, 변화의 축이라고 할 수 있는 사회적 모델의 강화 및 사회적 모델의 한계점에 대한 논의를 포괄하는 의미에서의 통합모델(개별모델과 사회적 모델의 통합적 관점, 또는 프레슬리의 통합모델)을 최근의 변화로 이해할 수 있음. 이에 대한 근거 논의는 이지수(2014)와 이종정(2018)의 논문에서 찾을 수 있음

인의 시민권 확보를 강조하면서 장애인을 억압하고 배제하는 사회적 요인의 변화를 요구하였다.

그러나 이러한 장애인권적 접근은 기존의 보호적 장애인복지와 성격이 다르기에, 서로 충돌하는 부분도 있으며 법적 제도적 차원은 물론 현실적 차원에서도 긴장과 공존의 관계를 이어가고 있기도 하다(전광석, 2011). 인권기반 복지실천의 등장에 관한 담론적 측면의 선행연구를 살펴보면, 장애인의 세력화와 함께 나타나는 지배담론과 저항 담론의 충돌 속에서의 기존 복지 수행 방식의 불가피한 전환이라고 평가되기도 하였고 (박선영, 2013), 장애패러다임의 변화 속에서 장애 관련 정책들이 변화된 내용을 수용하면서 나타나는 과도기적 부조화 현상이라고 해석하기도 하였다 (Waddington & Diller, 2002). 한편, 전광석(2011)은 수평적 패러다임을 지향하는 장애인권적 제도들과 수직적 보호적 패러다임을 지향하는 복지제도들은 상호변증적 보충관계라고 설명하면서 과도기적 상황은 아니고, 두 가지 다른 성격의 패러다임이 긴장과 공존, 보완적 성격을 유지하면서 공존할 것이라고 분석하였다. 과도기적 상황이라면 인권기반 접근이 보호적 접근을 대체해 나가야 하는데, 사회적으로 합의된 기준점까지의 욕구충족을 강조하는 복지의 전 영역을 대체할 수 없기 때문에 두 가지 논점은 공존할 것이라고 논의하였다.

이와 같은 인권 기반 복지실천의 가치철학에 대한 논의에는 정답은 없을 것이나 현재 장애인 복지현장에서 종사자들이 경험하고 있는 인권 기반 복지실천의 딜레마와 깊은 관련이 있는 것은 분명하다. 인권존중으로서 당사자의 자기결정권에 대한 무조건적인 인정을 할 것인가의 문제와 시설거주라는 단체생활에서의 통제나 보호, 질서유지를 위한 조치의 문제는 분명히 충돌할 수 있기 때문이다. 또한 안전이나 건강 등의 문제와 직결되는 이슈에서 장애인 당사자의 자기결정권과 종사자들의 규율이 갈등을 야기할 때 과연 인권 기반 복지실천이란 어떻게 하는 것인지 의문을 남긴다. 이와 같은 논의들은 종사자들이 인권감수성을 키워야 한다는 식으로 실천현장 속에서의 인권적 접근을 가능하게 하는 노력들을 강조하고 있으며, 장애인의 참여, 종사자와 장애인의 협력적 관계, 권리에 대한 옹호를 보장하는 방식에 대한 고민으로 이어지고 있다.

2. 장애인 거주시설과 인권적 복지실천

장애인 거주시설의 인권문제와 사회복지 실천방안과의 연계성에 초점을 둔 선행연구로는 김미옥 외(2008)의 연구가 첫 시도였다. 이후 후속연구들이 이루어졌으며, 지적 장애인 거주시설에서의 인권관점에 기초한 사회복지실천 경험에 관한 질적 사례 연구(김미옥·김경희, 2011)에서는 조직차원에서 인권보장위원회의 정비, 인권규정 제정, 인권교육 및 인권관점을 중시하는 조직문화 구축을 강조하였으며, 개인차원에서 이용자의 참여 증진, 종사자의 인권적 실천의 노력과 가치 변화, 인권침해상황에서 소통과 중재의 노력을 기울여야 한다고 논의하였다. 시설 내 종사자들이 경험하는 복지실천의 딜레마 연구가 전무한 상황에서, 다양한 논의를 이끌어낸 선도적 연구였다.

지적장애인 인권침해 사건 해결과정에서의 딜레마에 관한 질적 사례연구(김경희·김미옥, 2012)에서는 이용자와 종사자, 기관차원에서 각각 드러나는 인권 딜레마에 관한 연구를 수행하였는데, 이용자 차원에서는 지적장애 특성에 따른 자기진술능력, 자기행위 책임성과 관련된 딜레마가 도출되었고, 종사자 차원에서는 정당방어의 범위, 이용자 폭력에 대한 대처, 동료고발로 인한 갈등 등의 딜레마가 나타났다. 기관차원에서는 직원 대 이용자의 권리보장 충돌의 문제, 개인적 돌봄 대 공동생활의 갈등이 나타난다고 보고했다. 한 기관 내에서 발생한 하나의 사건을 다루었지만, 해결 과정에서 나타나는 다양한 딜레마를 분석한 것이 특징이다. 이와 같은 연구들은 다음 표2와 같이 정리된다.

〈표 2〉 장애인 거주시설에서의 딜레마와 복지실천에 관한 선행연구

저자명 (연도)	논문명	연구 방법 및 내용
김미옥, 정진경, 김희성 (2008)	장애인 거주시설의 인권 연구: 장애인당사자와 종사자의 인식을 중심으로	- (개별 및 포커스그룹 인터뷰) 장애인 거주시설의 인권문제와 사회복지 실천방안과의 연계성에 초점. - 장애인 집단과 종사자 집단을 구분하여 인터뷰하고 분석결과에서 인권보장의 영역별

		로 범주화하여 두 집단간의 인식이 차이를 보여줌
김미옥, 김경희 (2011)	인권관점에 기초한 사회복지실천 경험에 관한 질적 사례연구	- (질적 사례연구) 개인 차원에서 종사자들은 장애인의 자기결정권과 선택권을 중요하게 생각하는 동시에 실천과정에서 새로운 딜레마 경험하고 있음을 보여줌
김경희, 김미옥, 2012	지적장애인 인권 침해 사건 해결과정에서의 딜레마에 관한 질적 사례연구	- (질적 사례연구) 이용자 차원(지적장애 특성에 따른 자기진술능력, 자기행위 책임성과 관련된 딜레마 도출), 종사자 차원(정당방어의 범위, 이용자 폭력에 대한 대처, 동료고발로 인한 갈등 등의 딜레마), 기관차원(직원 대 이용자의 권리보장 충돌의 문제, 개인적 돌봄 대 공동생활의 갈등에서 드러나는 인권 딜레마)에서 딜레마를 분석함
김미옥, 김진우, 최영식, 윤덕찬. (2013)	거주시설 종사자의 딜레마 연구	- (질적 사례연구) 개인의 권리보장 대 공동생활의 문제, 자기결정권 존중 대 욕구파악의 어려움, 보호자의 욕구 대 기관의 책임 문제, 전문가로서의 역할과 가치 대 제한된 자원 문제, 안전 대 보호의 양상으로 딜레마를 분석함
박경수, 장혜경 (2013)	장애인 거주시설 실천가의 인권딜레마 사례연구	- (질적 사례연구) 갈등이 드러나는 세 가지 차원을 결과분석함. 개인차원에서는 안녕유지, 사적인 삶 보장, 의사표명 보장을, 공생차원에서는 타인의 안녕유지, 타인과의 적정한 관계유지, 사회규범의 유지를 드러냄. 물리적 제도적 구조의 차원에서는 물리적 환경의 보장, 장애인과 실무자의 관계유지, 인권민감성 보장의 주제를 제시함
조우연 외 (2013)	지적장애인 거주시설 사회복지사의 윤리적 딜레마 경험	- (Colaizzi의 현상학적 연구방법) 충남지역 지적장애인 거주시설에서 종사자들이 경험하는 실천의 딜레마를 조사한 결과 7개 범주 15개 주제로 나타남. '원 가장 복귀과정에서의 딜레마, 자기결정이 어려운 장애인들의 기본적인 욕구를 해결하는 과정에서의 딜레마' 등

거주시설 종사자의 딜레마 연구는 김미옥 외(2013)의 연구에 이르러 거시적 미시적 차원의 방대한 조사를 통해 종합적 체계적으로 분석되었다. 몇 가지로 결과를 정리하면, 개인의 권리보장 대 공동생활의 문제, 자기결정권 존중 대 욕구파악의 어려움, 보호자의 욕구 대 기관의 책임 문제, 전문가로서의 역할과 가치 대 제한된 자원 문제, 안전 대 보호의 양상으로 드러났다. 각 사례별로 해결방안까지 제시하고 있어서 실천적 함의가 큰 연구였다. 논의 부분에서는 실천의 딜레마가 나타나는 원인으로 전문직으로서의 윤리와 가치의 충돌, 한정된 자원의 분배, 전문실천기술개발의 미비 등을 지적하였다. 하지만, 해당 연구는 종사자들이 경험하는 상충되는 이슈중심으로 A 대 B의 딜레마적 경험에 주로 초점을 두었을 뿐, 딜레마를 포함하여 종사자가 경험하는 가치와 실천영역에서의 어려움과 혼란의 내용적인 측면이나 의미를 파악하는 데에는 한계가 있었다.

박경수·장혜경(2013)은 장애인 거주시설 실천가의 인권딜레마 사례연구를 수행하였는데, 인권관점에 기초한 복지실천을 할 때 갈등이 드러나는 지점들에 대하여 질적 연구를 수행하였다. 세 가지 차원으로 결과를 보고하였는데, 개인차원의 주제로는 안녕유지, 사적인 삶 보장, 의사표명 보장이 드러났고, 타인과의 공생차원의 주제로는 타인의 안녕유지, 타인과의 적정한 관계유지, 사회규범의 유지가 제시되었다. 물리적 제도적 구조의 차원에서는 물리적 환경의 보장, 장애인과 실무자의 관계유지, 인권민감성 보장의 주제를 제시하였다. 조우연 외(2013)는 지적장애인 거주시설에서 종사자들이 경험하는 실천의 딜레마를 조사하였는데, 원가정 복귀과정에서의 딜레마, 자기결정이 어려운 장애인들의 기본적인 욕구를 해결하는 과정에서의 딜레마, 장애인의 알권리와 비밀보장의 딜레마, 기관 내 구성원과의 가치관 차이에 대한 딜레마, 서비스 제공의 불평등으로 인한 딜레마, 지속되는 딜레마, 윤리적 극복을 위한 노력을 분석결과로 제시하였다. 이 연구들은 인권 딜레마 사례연구로서 어떤 상황에서 딜레마가 발생하는지에 관해 면밀히 분석했다는 점에서 가치가 있으나, 딜레마적 상황이 어떤 맥락에서 나타났는지에 관한 논의가 부족하고, 종사자들이 장애 인권에 대한 민감성을 키워야 한다거나 실천현장에서의 윤리적 지침서 제공이 필요함을 강조하고 있어서, 단기적 딜레마의 해결에만 초점을 둘 뿐 무엇이 인권적 접근인가와 관련한 대안 제시는 부족했다.

앞서 논의한 연구들은 사회복지실천의 영역에서 인권적 접근을 수행하기 위한 고민의 지점들을 드러내고, 복지의 큰 틀 속에서 이용자의 권리보장을 극대화 할 수 있는 방안을 찾고자 했다는 점에서 의의가 큰 것으로 평가된다. 하지만, 시설복지의 형태로써 최대한 인권적인 실천의 방식을 고민하는 연구들이었고, 변화하는 장애인복지 패러다임 속에서 거주시설 종사자들이 인권적 복지실천의 과정을 실제로 어떻게 경험하고 있고, 그 경험의 본질적 의미와 구성요소가 무엇인지는 알아보지 못하였다.

III. 연구방법

1. 연구설계

본 연구는 장애인거주시설 종사자들이 인권기반 복지실천을 수행경험이 어떠한지 알아보고 그 경험의 현상과 의미와 본질이 무엇인지 파악하고자 하였다. 이를 위해 Colaizzi의 현상학적 방법을 적용한 질적 연구를 수행하였으며, 그룹 내 역동을 통해 집중된 자료를 얻을 수 있는 포커스 그룹 인터뷰를 실시하였다. 현상학적 연구방법은 당사자의 입장에서 경험이나 인식의 의미를 확인하고 기술하는 연구방법인데(Giorgi, 1985), Colaizzi의 현상학적 연구방법은 인간 경험의 개인적 속성보다, 경험 전체의 공통된 속성이나 본질적 의미를 찾는 데 초점을 두는 연구방법이다(Colaizzi, 1978). 장애인거주시설 종사자들의 인권기반 복지실천의 경험속에서 공통된 속성을 파악하고 그 의미를 파악하는데 적절하다고 판단되었다. 또한 포커스 그룹 인터뷰는 특정 주제에 대한 인식이나 생각을 얻기 위한 토론식 자료수집 방법으로 알려져 있다(Morgan, 1996). 포커스 그룹 당 4-10명이 적합하며, 일반적인 소규모집단 토의와는 달리 특정 주제에 대하여 조직적인 논의를 진행하는 것으로, 적어도 3, 4개의 그룹으로부터 얻은 정보를 대조하여 질적으로 높은 자료를 수집 정리할 수 있는 방식이다(Krueger, 1988).

2. 연구 참여자

본 연구의 자료 수집은 공통된 경험의 범주를 한정하기 위해 의도적 표집(purposive sampling)을 통해 진행되었다. 경기도에 소재한 한 장애인 거주시설의 종사자 29명을 연구 참여자로 선정하였다. 이는 같은 환경에서 공통된 경험과 사건 속에서 인권적복지실천이라는 변화를 다양한 개별 종사자들이 어떻게 경험하고 있는지, 그 공통된 인식과 의미를 면밀하게 파악하기 위한 방안이기도 했다. 본 시설은 대규모 장애인 생활시설로서 현재 장애인 거주자들의 절반에 해당하는 인원이 탈시설을 시도하여 체험홈 등으로 독립해 간 상태이고, 절반정도의 인원이 남아있는 기관이었다. 장애인의 자기결정권과 자립생활을 강조하는 탈시설을 진행하고 있는 시설로서, 인권적복지실천의 경험을 연구하기에 적합한 시설이라고 판단되었다.

해당 거주시설의 종사자 중 거주시설 내에서 함께 생활하는 생활지도교사 전원과, 물리치료사, 영양사, 취사원, 사무국장까지 총 29명의 종사자를 면담하였다. 공통된 경험에 대한 집단의 역동을 활용하기 위하여 포커스 그룹을 활용한 심층 면담을 실시하였으며, 그룹 당 9-10명의 종사자가 참여하여 총 3개 집단에 대한 집단 면담이 이루어졌다. <표 2>에서 포커스 그룹 참여자에 대한 세부정보를 알 수 있으며, 참가자 대부분이 여성이었고, 평균 12년 4개월의 근속연수를 보이고 있어 해당 분야에서 오랫동안 근무한 이들로, 현재의 거주시설의 실태와 운영 등에 대한 의견까지 풍부한 자료수집이 가능했다. 해당 시설은 장애인 거주시설로서 인터뷰를 실시하기 몇 달 전 인권적 접근을 실천하고 있는지에 대하여 장애인 인권관련 외부 전문가들의 평가를 받은 바 있으며, 평가 이후에 종사자들은 인권적 접근을 해야 한다는 압박감으로 인해 어려움을 호소하고 있는 상황이었다. 특히 장애인 거주시설에서의 인권이라는 것과 관련하여 사회복지 실천을 어떻게 해나가야 하는가에 대한 고민이 많은 시점이었다. 한 개 시설에 대한 조사연구로서 연구의 제약점도 있겠으나, 이러한 목적적 표집(purposive sampling)을 통해 보다 면밀한 실태조사가 가능하였다.

〈표 2〉 포커스 그룹 구성원의 기본 정보

집단	번호	직종	성별	나이 (만)	근속연수
1	1	생활재활교사	여	42	11년 7개월
	2	생활재활교사	여	40	14년 3개월
	3	생활재활교사	여	50	15년 4개월
	4	생활재활교사	여	31	8년 5개월
	5	취사원	여	48	15년 6개월
	6	생활재활교사	여	23	3년 11개월
	7	생활재활교사	여	52	14년 9개월
	8	생활재활교사	여	43	21년 4개월
	9	간호사	여	48	17년 7개월
2	10	생활재활교사	여	41	2년 11개월
	11	생활재활교사	여	51	19년 10개월
	12	취사원	여	51	17년 4개월
	13	영양사	여	49	19년 8개월
	14	생활재활교사	여	24	13년 3개월
	15	물리치료사	남	35	7년 11개월
	16	사회재활교사	여	30	7년 9개월
	17	생활재활교사	여	33	3년 6개월
	18	간병인	여	60	13년 4개월
3	19	생활재활교사	여	42	14년 11개월
	20	사무국장	여	36	6년 9개월
	21	생활재활교사	여	36	2년 2개월
	22	생활재활교사	여	43	13년 3개월
	23	취사원	여	48	12년 5개월
	24	생활재활교사	여	51	13년 4개월
	25	생활재활교사	여	34	12년 1개월
	26	생활재활교사	여	30	13년 4개월
	27	생활재활교사	남	43	21년 5개월
	28	생활재활교사	여	53	10년 7개월
	29	생활재활교사	여	45	13년 4개월

3. 자료수집

본 연구의 자료수집은 2016년 8월에 한 그룹당 1회씩 시행하였으며, 1회당 인터뷰는 2시간 정도 소요되었다. 9명, 10명, 10명씩 한 그룹을 구성하여 포커스 면담을 진행하였으며, 그룹별 구성은 24시간 근무를 해야 하는 거주시설의 특성을 고려하여 비업무시간이 동일한 종사자별로 구성하였다. 한 기관에서 오래 근무한 이들이기에 친숙정도가 높아 집단의 역동을 활용한 인터뷰가 용이했다. 포커스 그룹을 활용한 면담은 이들에 걸쳐 진행되었으며, 8시간씩 3교대 하는 거주시설 내 종사자들이 참여하기 편한 시간인 업무 시작 전 또는 업무 종료후 시간으로 정하여, 시설 내 회의실에서 3개 그룹에 대하여 각각 진행하였다. 1차 인터뷰 분석결과를 통해 자료가 포화상태에 이를 때까지 전화로 추가 인터뷰를 진행하였고, 사무국장과의 추가 개별인터뷰를 통해서 기관에 대한 이해와 종사자의 업무배치 및 역할에 대한 이해도를 높였다. 인터뷰 내용은 생생한 기록으로 인터뷰 완료후 1주일 이내에 전사되었다.

포커스 그룹은 다음과 같은 질문순서로 진행되었고 반구조화된 양식으로 진행하여 답변을 들으면서 일부 조정되기도 하였다.

1. 사전질문: 각자 맡은 기관 내 직무를 포함하여 간단히 자기소개를 해주세요.
2. 도입질문: 장애인복지현장에서 인권적 복지실천이 강조되고 있는데 어떻게 실감하고 계시나요?
3. 주요질문: 인권적 복지실천이라고 하면 떠오르는 것은 무엇인지 말씀해주세요.
4. 주요질문: 인권적복지실천의 경험을 말씀해주세요. 어려웠던 점이나 실천 사례 등 느끼거나 생각한 것 등 자유롭게 이야기해주세요.
5. 주요질문: 인권적 복지실천을 위해 현장에서 정말 필요로 하는 것은 무엇인지 말씀해주세요.
6. 마무리: 말씀하신 내용을 정리하겠습니다. 또한 이외에도 추가로 궁금하거나 하실 이야기가 있으면 해주세요

4. 윤리적 고려

자료수집에 앞서 주제 개념에 대한 민감성을 높이기 위하여, 연구자는 연구 주제와 관련한 연구 보고서 및 학술지 연구물들을 참고하였다. 또한 관련 선행연구의 검토와 거주시설에서의 봉사경험을 토대로 기존의 시설종사자들의 경험에 대한 감수성을 높이고자 노력하였다.

연구의 윤리성과 연구참여자의 비밀보장 및 익명성을 확보하기 위해 연구동의를 구하였고, 연구 참여에 대한 자발성을 확인하고 언제든지 연구에 참여를 중지할 수 있음을 공유하였다. 개방형 질문으로 논의를 시작하였으며, 자연스럽게 집단 구성원들의 의견이 오갈 수 있도록 하였다. 연구자는 포커스 그룹 진행 전에 미리 질문과 진행방식을 계획하였으며, 진행시에는 참여자들의 동의를 얻어 오디오로 기록하였고, 포커스 그룹 진행자 외에 기록을 담당하도록 장애관련 박사과정 대학원생 1인을 참가하도록 하였다. 인터뷰 기록의 목적인 연구에만 활용될 것임을 분명히 설명하였고, 논의 주제가 무엇인지 구체적으로 설명하였다. 수집된 자료는 연구자만 열람하며 연구 종료 후에는 폐기될 것을 약속하였다.

5. 분석방법

장애인 거주시설 종사자의 인권기반 복지실천 경험 속 본질과 의미를 찾기 위해서, 현상학적 분석방법인 Colaizzi(1978)의 자료분석방식을 적용하였다. 첫째, 녹음자료를 그대로 전사하였고, 전사한 내용을 반복적으로 읽으면서 연구주제와 관련하여 의미 있는 진술을 추출하였다. 둘째, 의미있는 진술을 여러번 읽으면서 조사하는 현상과 관련된 단어, 문장, 구절을 찾고 맥락을 고려하여 연구자의 말과 언어로 재구성하였다. 셋째, 구성된 의미에서 유사한 것끼리 묶어서 주제를 범주화 하였다. 그 과정에서 종사자들의 경험과 맥락이 일치하는지 범주화과정에서 모순은 없는지 원자료와 비교하여 확인하였다. 넷째, 연구참여자들의 느낌과 반응에 초점을 두고 주제가 상하위 속성 간에 맥락적으로 일치하는지, 주제들 간에 차별화가 되는지 확인하였다. 다섯째,

분석된 자료를 주제와 범주에 따라서 나열, 해석, 기술하였다. 경험의 의미를 통합적으로 이해하며 기술하고자 노력하였으며, 연구의 신뢰도를 확보하기 위해 포커스 그룹 인터뷰에 기록을 하며 참관했던 박사과정 대학원생에게 범주의 분류에서 과장되거나 과소하게 분석된 부분이 없는지, 유의미한 발언들 중 배제된 내용은 없는지 검토하도록 하였다. 또한 분석결과를 연구 참여자들에게 보내어 본인들의 경험에 대한 해석이나 분류가 적절하게 이루어졌는지 검토하도록 하였다.

6. 연구의 엄격성

Lincoln과 Guba(1985)는 연구의 엄격한 수행을 위해 ‘사실적 가치’, ‘적용성’, ‘일관성’, ‘중립성’ 면에서 연구의 신뢰도와 타당도가 확보되어야 함을 강조하였다. 사실적 가치는 연구가 얼마나 현실을 잘 반영하고 있는가를 의미한다. 본 연구에서는 분석한 연구결과가 얼마나 종사자의 경험을 잘 반영하고 있는가를 확인하기 위해 연구자가 반복 검토하였고 인터뷰 전일정에 참관한 박사과정생의 검토, 연구참여자의 최종검토를 거쳤다. 적용성은 연구 결과가 다양한 상황과 맥락에서도 적용될 수 있는가를 의미한다. 인권적복지실천이라는 변화를 경험하는 다른 거주시설이나 장애인복지현장에도 충분히 적용할 수 있도록 연구 대상 기관의 종사자 대부분을 연구참여자로 확보하였고, 기관 및 종사자 개인의 상황과 업무에 대하여 연구자가 충분한 이해를 하고자 하였다. 또한 참여자의 진술이 포화상태에 이를 때까지 추가 인터뷰를 통해 진술을 확보하였다. 일관성은 연구분석결과와 신뢰도를 의미한다. 발견된 주제에 대하여 공통적으로 나타나는 주제인지 지속적 비교를 시도하였다. 중립성은 편견을 배제하는 것을 의미한다. 분석시 연구자의 편견을 배제하고자 노력하였으며, 관련분야 전문가의 검토를 받았다.

IV. 연구결과

아래의 표에서 제시하고 있듯이, 장애인 거주시설 종사자들이 경험하는 인권기반

복지실천의 어려움은 3가지 상위범주로 구분된다. ‘실존적 가치의 혼란’과 ‘철학적 가치의 혼란’, ‘실천적 방향상의 혼란’을 경험하고 있었다.

〈표 3〉 분석결과 범주화

상위범주	하위범주	속성
실존적 가치의 혼란	종사자로서의 목소리 상실	역차별을 경험, 장애가 특권이 됨, 종사자의 처우나 입장을 주장할 곳이 전혀 없음, 일자리도 없어 질까 걱정
	존재 가치의 상실	그간의 헌신과 희생에 대하여 인정받지 못함, 외부의 이상한 시선, 희생자에서 억압자가 되는 경험
철학적 가치의 혼란	자립생활 철학에 대한 의문	장애인의 완전한 자립 가능성에 대한 의문, 탈시설과 지역사회보호의 필요성에 대한 의문
	자기결정권의 본질에 대한 의문	장애인의 잘못된 선택에 대해서도 존중해야 하는지의 의문, 교육과 훈육이 필요한 시설 거주 장애 청소년의 자기결정권의 인정 문제
실천적 방향상의 혼란	자기방어적 실천	실천 시 면책 장치 마련이 우선이 됨, 소극적 복지 프로그램의 실천, 방임적 관리자적 태도
	기존 실천방식에 대한 고민	이용자 간 갈등 문제, 복지사와 장애인의 관계와 호칭의 문제, 개인 기호도 존중 수준의 문제, 동성애나 피임 방식 등에 대한 교육문제, 상호 간 인권감수성 충돌의 문제 등

장애인 당사자들의 인권과 자기 결정권이 강조되면서, 장애인들의 생활 전반에 걸쳐 장애인 당사자의 자기결정권이 강조되는 한편 종사자들의 목소리는 상실되었으며, 이와 더불어 존재 가치의 상실도 경험하고 있었다. 장애인 당사자의 자립 생활과 자기결정권에 대한 의문을 표명하는 등, 인권기반 복지접근으로의 장애정책의 방향 전환에 대하여 기존에 수행하던 실천방식의 기저에 깔려있던 복지철학에 대한 혼란도 경험하고 있었다. 철학적 측면뿐만 아니라 실천적인 차원에서도 기존의 헌신적 접근과는 사뭇 다른 자기방어적인 복지실천을 하고 있거나, 기존에 해왔던 복지실천 방식과 관련하여 다양한 이슈를 중심으로 실천방법을 고민하는 모습도 나타났다.

1. 실존적 가치의 혼란

가. 종사자로서의 목소리 상실

종사자들은 인권적 접근이 강조되면서 역차별을 경험하며, 장애가 특권이 되는 상황으로까지 느끼고 있었다. 과거에는 부모 같은 마음으로 장애인 이용자들에 대하여 사랑과 보호라는 취지하에서 그들의 목소리가 존중받았다면, 인권적 접근으로 인하여 당사자의 자기결정 권리가 강해지고, 인권접근에 대한 강력한 정책적 요구와 인권침해 조사 등이 이루어지다 보니, 복지사를 비롯한 시설 내 치료사 영양사 등 전체 종사자들의 목소리가 줄어들고 있었다. 이런 상황에서 종사자들은 자신의 처우나 입장을 주장할 곳이 전혀 없다고 하면서, 억울함을 호소하기도 했다. 이러한 자신들의 목소리의 상실은 장애인의 돌발 및 문제행동인 도전적 행동으로 인한 부상을 당할 때에도 종사자로서 자신을 권리를 주장할 수 없게 되기까지 하였다. 또한 종사자의 목소리의 상실은 점차 자신들이 설 자리가 없어질지도 모른다는 일자리 상실에 대한 두려움으로 나타나기도 했다.

현재 종사자의 인권이 보장되지 않아서 종사자에게 피해가 돌아가는 일이 많아요. 인권조사를 나왔을 때 학생이 도둑질을 한 사건에 대해서도, 그 도둑질을 문제 삼는 게 아니라 학생의 몸을 뒤흔 선생님이 문제가 되요. (참여자 8)

이용자들 기호에 따라서 식단을 운영했는데, 인권조사에서 영양을 고려해서 잡곡밥으로 하라고 지적을 했어요. 그래서 잡곡밥으로 하고 있는데, 근데 일부 아이들은 잡곡밥 싫다고 울고 있고, 선생님들이 밥 먹는 것까지 제재한다고 불만을 토로하는 경우가 많아요. 종사자 입장에서 뭐라고 한 마디 할 수도 없어요. (참여자 3)

어떤 물건을 사달라고 했을 때 안사주면 짜증내고 담당 선생님을 바꾸겠다고 하는 학생도 있어요. 선생님이 금전관리에 관여하면 인권을 억압했다고 고발한다는 학생도 있고요. (참여자 1)

이용자(장애인)들의 인권에 대해서는 많이 강조하고 있는데, 우리 종사자들의 인권에 대해서는 어디 가서 얘기도 못 꺼내요. (참여자 21)

부모님이 외부 프로그램을 원하지 않을 때, 우리는 부모님의 의견을 따르게 되는데요. 이용자들은 우리들이 나가지 못하게 한다고 인권조사관한테 고발을 해요. 결국 우리만 피해를 보게 되는 거죠. (참여자 14)

그냥 빨리 다 탈시설했으면 좋겠어요. 그러면 여기서 이렇게 아옹다옹할 일도 없을 거고요. 근데, 탈시설하면 우리는 어디서 일하게 될까하는 걱정도 되요. (참여자 11)

나. 존재가치의 상실

시설 종사자들은 그들의 존재가치의 상실도 경험하고 있었다. 평균적으로 십여 년이 넘도록 시설근무를 해 올 만큼, 이 일이 자신의 삶의 일부가 되어있고, 0세 때부터 시설에서 생활해 온 이용자들과는 가족과 같은 생각도 있었는데, 인권적 접근을 시도하면서 그간의 헌신과 희생에 대하여 인정받지 못하게 되어 힘들어하고 있었다. 과거에는 시설에서 일하는 복지 종사자들에 대하여, 좋은 일 한다고 헌신의 상징으로 대했다면, 요즘에는 장애인을 억압할지도 모르는 사람으로 대한다고 느끼고 있었다. 복지 종사자로서의 높게 평가되던 존재 가치를 잃어가고 있는 것이다.

장애인 시설에서 근무한다고 하면, 이상하게 봐요. 언론에서 시설 장애인들이 착취당하거나 인권유린을 당하는 내용이 많이 보도되어서 그런지, 우리를 이상하게 봐요. (참여자 19)

정신과 약을 먹는 학생과 병원에 갔을 때, 의사는 장애인 학생의 말만 들어주려고 하고, 함께 간 시설 선생님의 얘기는 듣지 않아요. 시설 종사자의 말을 믿지도 않거나 뭐가 신뢰할 수 없는 사람으로 취급받는 기분이예요. (참여자 17)

인권조사 나오는 분들은 우리를 막 대해요. 이용자 분들이 인권침해라고 이야기 하는 부분은 전부 다 인정하고, 벽보고 서있으라고 중2인 이용자에게 체벌

한 것도 걸리니까, 난 여기서 뭐하는 사람인가 싶고요. (참여자 2)

2. 철학적 가치의 혼란

가. 자립생활 철학에 대한 의문

장애인의 자립 생활을 강조하면서 탈시설화 작업이 진행되었고, 시설 내 남아있는 장애인들에 대해서는 인권기반 복지실천이 강조되기 시작했다. 그러나 시설종사자들은 장애인의 완전한 자립 가능성에 대한 의문을 제기하면서, 종종 장애인의 자립 생활이 이념적으로는 옳으나 현실적으로 어려우며, 탈시설과 지역사회케어의 방향성에 대한 의문을 제기하고 있었다. 특히 장애인들이 자립 생활을 주장하고 있으나, 권리만 강조하고 있다는 점과, 경쟁적인 사회 환경에서 노동을 통한 자립의지도 부족하다는 점, 일상 속에서 종사자들에 대한 의존도가 높은 상황에서 자립이 가능한가에 대한 의문, 본질적으로는 탈시설과 자립 생활의 철학이 과연 현재의 장애인들에게 적합한 장애정책의 방향인지 의문을 제기하고 있었다.

자립 생활은 스스로의 삶에서 자기결정권을 갖는 거라고 하잖아요. 그런데, 이 사용자들 중에는 받을 건 받고 성인으로서 자기 의무는 다하지 않는 경우가 있어요. 장애인의 권리만 강조하고 의무는 이행하지 않는 경우, 이걸 불공평한 것 같아요. (참여자 15)

장애인도 직장 다니고 탈수급을 해서 떳떳하게 살아갔으면 하지만, 내가 볼 때 장애인들은 노동을 통해 독립하고 싶다는 의지는 약하고, 구체적인 계획은 없는 편이더라고요. (참여자 27)

이념적으로는 장애인들이 자립생활을 할 수 있다고 하는데, 사실 우리 주변을 보면 우리 종사자들에게 많이 의지하고 선생님들이 다 해줄 거라고 생각하는 경우가 많아요. 선생님들이 일상적인건 다 해주고 그냥 장애인 이용자들은 직장만 왔다갔다 하는거죠. (참여자 29)

탈시설 했다가 다시 시설로 복귀하는 경우가 많아요. (중략) 시설장애인들은 사회에 나가면 참아야 하는데 (중략) 나가도 적응도 못하고 다시 돌아오죠. (참여자 22)

시설에서는 우선 기본적으로 먹고사는 문제가 해결이 되니까, 머리가 좀 되는 이용자는 세상이 경쟁이 심하다는 것을 아니까 안 나가려고 하고, 오히려 세상을 잘 모르는 이용자들이 나갔다가 다시 돌아오는 경우가 많아요. (참여자 25)

탈시설을 주장하는 활동가들도 선별된 장애인만 탈시설이 가능하다는 걸 알아요. 탈시설이나 자립이야기를 많이 하는데 연구자 선생님은 전체 장애인이 다 실제로 자립이 될 거라고 생각하세요?. (참여자 16)

나. 자기결정권의 본질에 대한 의문

자기결정권에 대한 존중과 관련하여 장애인의 잘못된 선택에 대해서도 종사자들은 존중해야 하는지 의문이 든다고 하였다. 특히 교육과 훈육이 필요한 시설 거주 장애 청소년의 자기결정권의 인정 문제가, 청소년 이용자들의 비행과 연결될 때 어느 정도 수위에서 강하게 교육을 할 수 있을지 의문을 제기하고 있었다. 위생관리나 건강관리 등 꼭 수행해야 하는 일에 대해서도 이용자들이 자기결정권을 주장하고 있을 때 종사자들과 의견충돌을 빚고 있었으며 동성 간 스킨십과 같은 문제에 대해서도 사회적인 일반적 룰인 이성애에 대한 교육을 강조해야 하는지도 의문을 제기하고 있었다.

중학교 3학년인 학생인데 돈 씹셈이가 너무 크고 장애인 수당으로 나오는 돈 보다 몇 배 더 많은 지출을 해요. 선생님으로서 미래준비를 위해 아껴쓰라고 해도 별로 효과도 없어요. 현금지출을 못하게 하면, 교통카드를 충전해서 편의점에서 다 쓰고 자기 용돈의 몇 배가 되는 액수의 선물을 사서 반 친구들에게 다 나누어주고, 돈이 없으면 외부 상점에서 물건을 훔치기도 해요. (참여자 2)

개인위생관리가 전혀 안 되는 이용자가 있어요. 담임선생님이 지적을 하고 갈아입을 옷을 가져다주어도 갈아입지 않아요. 선생님과 눈도 안 마주치려 하고 대화도 피해요. 자기 결정으로 존중하기에는 냄새도 나고... 심하게 지적을 하

면 스트레스 받아서 경기를 하거든요. (참여자 8)

시설은 동성을 접하기 쉬운 상황인데, 동성 간의 스킨십 문제를 놔두어야 하는지 모르겠어요. 과도하게 특정 이용자 들어서 스킨십을 하는 경우에도 자기결정권 존중으로 그냥 허용해야 하는지 모르겠어요. (참여자 29)

벨라톤으로 소변을 빼야 하는 이용자가 있는데, 소변 빼라고 하면 선생님한테 짜증내고 소리를 질러요. 잘 때도 안 하고 그냥 자기도 하고, 주말에는 아침에 소변 한번 빼고 다시 자라고 하는데, 이용자는 이걸 잔소리로 들어요. 자기결정권이라고 존중하기에는 안 좋은 일이지요. (참여자 9)

3. 실천적 방향상의 혼란

가. 자기방어적 실천

거주시설의 종사자들은 인권기반 복지실천을 이행하기 위해 자기 책임 회피장치 마련을 먼저하고 있었다. 인권접근이 강조되고, 시설평가에도 얼마나 자기결정권이 존중되고 있는지를 중요시하고, 인권조사관이 나와서 인권침해 조사를 하고 있기에, 복지 프로그램 하나하나 실천하는 데 있어서 이용자들의 권리를 존중하고 있다는 증빙서류를 만들어야 하는 것이다. 이로 인하여 소극적 복지 프로그램을 실천하고 있었으며, 사고 없이 하루하루 지나가기만을 바라는 등 방임적 태도를 보이고도 있었다. 특히 장애 청소년을 대상으로 지원하고 있는 종사자들의 이러한 어려움을 가장 큰 어려움으로 호소하고 있었다.

인권적 접근이 강조되면서 장애 청소년 애들한테 체벌을 할 수도 없고, 제대로 훈육을 할 수가 없어요. 이용자들 중에서 “그냥 별점 매기세요”하고 가버리기도 하고요. 이젠 연루되고 싶지가 않아요. 애들끼리 싸우든지 말든지 그냥 관여하고 싶지가 않아요. (참여자 1)

애들은 학교 다녀오면 운동이고 공부고 다 싫고 컴퓨터만 하려고 해요. 그만하라고 하기도 나도 힘들고, 또 애들이 컴퓨터를 안 하면 할 일이 없어서 서로 더 많이 싸우고 난리가 나요. 그러니까 그냥 컴퓨터 하게 하면서, 사고나 다름없이 하루하루 보내는 데에만 신경을 쓰게 돼요. (참여자 19)

연고가 있는 이용자들의 경우에는, 무슨 활동을 하든지 연고자에게 허락을 구한 후에 해요. 동의서 같은 것을 받아둘 수 있는 거죠. 맨날 나를 보호하기 위한 동의서 같은 서류 작업하다 보면 내가 장애인들을 위하는 일을 하는 게 맞나 그런 생각도 들어요. (참여자 21)

나. 기존의 실천 방식에 대한 고민

인권기반 복지실천 방식의 모호함을 여러 가지 실질적 이슈에서 제기하고 있었다. 이용자 간 갈등의 문제, 복지사와 장애인의 관계와 호칭의 문제, 개인 기호도 존중 수준의 문제, 피임 방식 등에 대한 교육문제, 상호 간 인권감수성 충돌의 문제 등에 대해서, 기존에 하던 것과 달리 어떻게 하는 것이 인권기반 복지실천인지에 대한 실천적 의문들을 가지고 있었다.

방 배치를 하다 보면 어떤 이용자는 방에서 막 소리 지르고 울고 하는데, 같은 방 사람들은 시끄럽다고 하고 선생님 입장에서는 어떻게 해야 할지 모르겠어요. 교육하고 통제하기 힘든 이용자들을 어떻게 관리하고 교육해야 인권의 문제가 안 되는 걸까요? (참여자 24)

외부 직장을 다니면서 남자친구가 있는 이용자가 있어요. 회사 다녀올 때면, 피임약을 습관처럼 오자마자 먹도록 하고 있는데 이는 인권침해 아닌가요? 무연고자라서 상의할 데도 없고, 어떻게 해야 할지 모르겠어요. (참여자 12)

외부로 학교를 다니는 여학생들에 대해 임신 가능성이 있어서, 매달 생리체를 하고 있거든요. 직접적인 체크가 아니더라도 생리대에 번호를 매기고 현재 몇 개 썼는지 체크하는 식으로 하면, 인권적인 걸까요? 이런 체크가 필요한 걸까요? (참여자 10)

이용자들이랑 장난삼아 이야기 할 때, 어느 정도 수위를 지켜야 하는지 고민이 되요. 농담 삼아 던진 말이 인권침해는 아닌지 말이죠. 이용자들과 오랜 세월 함께 하다 보니, 이름이 아닌 별명을 부르게 되요. 별명을 부르는 것이 문제가 되는지도 궁금해요. (참여자 14)

V. 논의 및 결론

인권기반 복지실천이 강조되면서 시설현장에서 종사자들의 경험은 ‘다양한 혼란상태의 경험’으로 축약될 수 있었다. 논문 결과에서 분석한 대로 시설종사자들은 인권기반복지실천을 하는 가운데서 다양한 혼란을 경험하고 있는 것으로 나타났으며, 결과를 통해 파악한 종사자들이 경험한 혼란의 본질적 양상은 ‘자신의 실존적 가치의 혼란, 장애인 자립 및 자기결정권에 대한 철학적 가치의 혼란, 복지실천방식의 혼란’으로 정리되었다. 그 중에서도 자립 생활의 핵심 철학인 자기결정권에 대한 존중이 종사자들의 자기방어적 복지실천을 초래하고 있었는데, 이는 인권적 실천이라는 철학적 이상과 현실 사이의 괴리를 여실히 보여주었다. 그 외의 연구결과를 토대로 다음의 몇 가지 논의가 가능하다.

첫째, 연구결과를 통해 볼 때, 인권기반 복지실천이라는 장애정책상 패러다임의 전환이 다양한 딜레마적 상황들을 만들고 있었으며, 시설종사자들은 이러한 상황에서 실존적 가치의 혼란, 철학적 가치의 혼란, 실천 방향의 혼란을 경험할 수밖에 없었다. 기존의 재활과 보호 중심의 접근과 인권기반 장애인복지실천의 접근은 이념적으로도 상충할 뿐만 아니라, 현실세계에서는 세부적인 일상의 갈등사건으로 드러나고 있었다. 결국 이러한 어려움을 야기한 것은 인권기반 복지실천으로의 장애인복지 정책의 기조 변화였으며, 개념과 시도 자체가 두 가지 상충적인 내용을 담고 있어 실천현장에서 종사자에게 내적 혼란을 야기하였고, 이용자와의 관계에서는 일상에서의 갈등사건을 야기하는 것으로 해석된다. 하지만 이를 장애인복지철학이 인권중심으로 무게중심을 옮겨가는 과도기적 현상으로 해석하고 그 안에서의 다른 담론 또는 철학이 실천양상에서의 충돌로 나타난 것이라고만 해석하기에는 무리가 있다.

그렇다면 이 현상을 과도기가 아닌 어떤 현상이자 의미로 해석할 수 있을까? 과도기라면 점점 인권중심기반으로 옮겨가면서 보호적 기능이 사라져야 함에도 불구하고, 현실에서는 다양한 다른 형태의 장애인에 대한 보호적 기능을 하는 공간이 생겨나고 있다. 예를들면 대형 거주시설은 30인 이하 거주시설로의 전환이 요구되는 법적 제재 속에서 소규모화 되고 있고, 그룹홈이 늘어나고 있다. 즉 거주시설내의 인권적 복지실천의 공간의 재구성이 그룹홈의 형태로 진행된다고 볼 수도 있다. 이는 복지와 인권의 절충점이자, 장애인의 지역에서의 삶과 자기결정권을 강조하면서 보호체계 속에 안전하게 살아가게 하는 현실적 복지제공 실천의 타협점이라고 본다. 즉, 재활과 보호 중심의 기존 장애인 복지 패러다임과 인권에 기반을 둔 새롭게 강조되는 장애인 복지 패러다임이 ('상충'하는 것이 아니라) '공존'할 수 있는 방안을 찾아가고 있는 것이다. 전광석(2011)도 복지란 사회적 합의의 수준에서 이행되는 욕구충족이기 때문에 인권적 복지실천이 강조된다고 하더라도 기존의 보호적 복지실천이 공존할 수 밖에 없음을 논하였다. 장애인복지 실천현장에서 나타나는 다양한 혼란과 어려움은 장애인복지 패러다임의 전환속에서 새로운 인권적 복지실천의 방향과 기존의 관점 사이의 혼란 양상으로 나타나고 있으나, 패러다임이 인권적 복지실천으로 완전히 전환하게 되기 보다는 '보호와 인권' 두 가지를 모두 절충하는 타협점(이용자와 종사자의 합의의 수준)을 찾는 방식으로 해결될 여지가 많다고 생각한다. '장애인의 자기결정권을 존중하는 전문가주의'라는 복지체계와 실천방식을 만들어나갈 것으로 전망한다.

이미 그러한 노력들이 현장에서 진행되었다. 최근 몇 년간 거주시설 종사자를 대상으로 다양한 보수교육이 실시되었고, 시설협회 등에서 인권적 복지실천을 위한 매뉴얼을 개발하여 보급하는 등의 노력을 기울였다. 장애인의 도전적 행동에 대한 인도주의적 접근 매뉴얼도 개발되는 등 실천적 전략들도 보급되고 있다. 그럼에도 이런 현상이 나타나는 것은 본질적으로 시설종사자들이 자립이나 자기 결정에 관한 가치철학에 대해서도 충분히 진심으로 동의 되지 않았기 때문일 것이다. 방법론적인 측면에서 장애인의 인권에 기반한 복지실천방안을 마련하는 것도 중요하겠으나, 본질적으로 '장애인을 자립과 자기결정권을 가진 독립적 존재'로 바라보고, 종사자의 역할이 조력자이자 지역사회로 장애인이 나아가고 활동하게 하는 촉진자에 있음을 깊이 있게 공감하고 받아들일 수 있도록 하는 노력이 더 필요하다. 보호적 관점과 인권적 관점의

공존이 지속된다고 하더라도 보다 이용자의 입장에서 더 좋은 절충점(혼란 또는 딜레마의 해결책)을 찾아야 할 것이다.

둘째, 장애패러다임의 전환으로 인한 복지인력들이 경험하는 혼란과 갈등 양상이 장애인 생활시설에만 국한되지는 않는다. 특수교육의 현장이나 보호작업장, 복지관 등에서 일하는 기존의 복지제공인력들도 유사한 경험을 하고 있으며, 장애인식개선교육이 전 국민을 대상으로 이루어지고 있음에도 불구하고 장애인의 지역 내 통합을 둘러싼 한국사회의 고정관념과 편견의 장벽은 여전히 존재한다. 그렇기에 장애패러다임 전환으로 인한 현장에서의 변화와 어려움에 관한 후속연구들은 계속 필요할 것이며, 복지제공 인력의 인권기반 복지철학 정립 및 실천방식 기준 마련이 필요할 것이다.

셋째, 당사자의 인권도 중요하지만 종사자의 인권도 중요하게 여겨져야 한다. 그동안 억압받고 차별받은 장애인들의 인권에 대한 보장이 사회적 이슈가 되면서, 사회복지 영역은 보호에서 인권으로 패러다임의 전환을 경험하고 있다. 이러한 과도기 속에서 종사자는 매일의 일상에서 딜레마에 빠지며 어찌할 바를 모르는 상황에 놓인다. 종사자 역시도 역사의 과도기 속에서 발생할 수밖에 없는, 한국 장애인에 대한 억압과 해방의 역사가 낳은 불가피한 결과일 수도 있다. 아마도 시간이 지나면서 균형점을 찾아갈 것이다. 인권기반 복지실천이라는 딜레마에 빠질 수밖에 없는 시설이라는 공간이 우리 사회에 불가피한 사회제도의 형태로 지속된다면, 종사자와 이용자가 서로를 이해하고, 서로의 인권을 존중하는 균형점을 반드시 찾아야 한다. 이를 위해서는 종사자를 대상으로 인권적 복지실천을 위한 교육을 실시하는 것과 더불어, 장애인 당사자를 대상으로 하는 복지제공자와의 인권적 관계증진을 위한 방안을 교육할 필요가 있다. 특히 장애인 당사자의 인권과 의무, 또는 책임이 무엇인지, 종사자와 당사자 사이에서의 인권과 교육의 경계, 종사자와 이용자 간의 서로에 대한 예의 등에 관련하여 중점적인 논의가 필요하다. 실천현장에서는 딜레마적 상황에 대한 타개책으로 시설 내부에서 종사자와 이용자 간의 집단토의 시간을 통해 구체적인 상황에 대한 실천수칙을 함께 정하는 등의 대안 모색도 가능할 것이다.

넷째, 기존연구에서 제안하고 있는 종사자에 대한 인권교육이나 가이드라인이 인권적 복지실천을 위한 도구는 될 수 있으나, 종사자와 이용자를 위한 진정한 대안이 되기에는 한계가 존재한다. 연구결과에서 드러났듯이 매일의 일상 속에서 복지와 인권

의 딜레마는 종사자들에게 큰 문제가 되고 있다. 특히 실존 가치의 딜레마와 철학적 가치의 딜레마는 종사자의 존재 가치 자체에 대한 의문을 야기하고 있기에 단순히 ‘인권적 접근의 가이드라인’ 제공만으로 해결할 수는 없는 문제인 것이다. 물론 기존의 연구들은 딜레마적 상황의 해결을 위해 인권감수성을 키워야 한다거나, 인권접근을 위한 교육 등을 대안으로 제시하고 있다. 또한 전문직으로서의 윤리와 가치의 충돌, 한정된 자원의 분배, 전문실천기술개발의 미비 등을 딜레마의 원인으로 언급하면서 인권적 실천기술개발을 강조하고 있다(김미옥 외, 2013). 그러나 장애패러다임의 변화 속에서 종사자들이 ‘내가 여기서 무얼하고 있나’ 하는 실존 의미상 딜레마를 느끼는 현실은, 전문직으로서의 실천적 방법상의 가치 충돌의 문제만은 아닌 것임을 보여준다. 종사자와 이용자 사이의 이해 부족이나 서로의 인권보장에 대한 노력이 필요함을 간과하고 일방적으로 종사자에게만 인권적 실천을 하라고 하는 현실 때문일 수 있다. 이에 앞서 제시한 대로 장애패러다임 변화에 대한 깊이 있는 종사자들의 이해가 선행되어야 할 것이며, 장애인과 종사자가 서로 의존해있는 공생관계임을 알고, 향후 지역 사회 내 장애인의 자립을 적극적으로 시설종사자들이 나서서 도울 수 있는 방안을 고민해야 할 것이다. 그러할 때 비로소 본인들의 역할 가치도 다시 찾을 수 있을 것이며, 향후 일자리에 대한 위기감도 사라질 것이고, 인권기반 복지실천의 틀을 만들어 나가면서 상호 간의 인권을 존중하는 전략을 만들어낼 수 있을 것이다.

본 연구는 다음의 한계점을 지닌다. 첫째, 한 기관을 대상으로 조사를 수행했기에 논의를 끌어내는 데에 있어 다소 기관의 특수성이 반영되었을 수 있다. 둘째, 주로 종사자의 입장을 중심으로 조사를 수행하였기에 장애인 당사자의 관점이 없어서 역시 풍부한 논의를 끌어내는 데에는 한계가 존재한다. 그러나 다양한 직종의 종사자를 포함하였고, 종사자가 경험하는 어려움의 본질을 해석하는 데 있어서 연구 참여자의 관점을 최대한 반영하고자 노력하였다. 셋째, 연구동향을 살피면서 해외자료를 충분히 검토하지 못한 점도 한계이다. 패러다임의 전환이 국제적 현상인 만큼 추후 연구에서는 국내외 인권적복지실천의 과정 속에서 종사자의 경험의 차이 등 비교연구가 수행된다면 국내적 차원의 문제해결 방향을 논하는데 도움이 될 것이다. 넷째, 본 연구는 질적 연구로 수행되면서, Colaizzi의 현상학적 연구방법인 “인간 경험의 개인적 속성보다, 경험 전체의 공통된 속성이나 본질적 의미를 찾는 데 초점을 두는 연구방

법”을 적용하고자 하였음에도, 연구결과에서는 다소 ‘경험의 개인적 속성’으로 읽혀질 수 있는 부분이 있었다. 기존의 개인경험에 대한 포커스 그룹연구와의 차별성이 드러나지 않은 것이 한계라 할 수 있다. 균질한 집단에서 일하는 종사자였기에 공통적 경험으로서 해석하였다고 하더라도, 인권적 복지실천의 경험은 개인적 속성의 차원에서 해석된 것이다. 후속연구에서는 보다 명확하고 세심한 Colaizzi의 현상학적 연구방법이 적용될 필요가 있다. 그럼에도 본 연구는 장애인복지 패러다임에 대한 논의에서부터 실천현장의 상황을 포괄하는 연구로서, 장애인 거주시설 종사자의 인권기반 복지실천을 수행하는데 있어서의 경험과 의미를 파악하는 데에 관심을 두었다는 점에서 가치 있다. 딜레마적 이분법적 상황에 초점을 둔 선행연구와 구별되며, 포괄적 의미에서의 가치와 실천, 철학에서 경험하는 혼란의 내용과 의미를 살펴본 점에서 의의가 있다.

참고문헌

- 김경희·김미옥. 2012. “지적장애인 인권침해사건 해결과정에서의 딜레마에 관한 질적 사례연구.” 『한국장애인복지학』 제19권: 235-252.
- 김미옥·정진경·김희성. 2008. “장애인 거주시설의 인권 연구: 장애인당사자와 종사자의 인식을 중심으로.” 『사회복지정책』 제33권: 389-422.
- 김미옥·김경희. 2011. “인권관점에 기초한 사회복지실천 경험에 관한 질적 사례연구.” 『한국사회복지학』 제63권 1호: 29-55.
- 김미옥·김진우·최영식·윤덕찬. 2013. 『장애인 거주시설의 서비스 딜레마(문제점) 사례연구』 한국장애인복지시설협회 수탁연구.
- 김선주. 2016. “장애인시설종사자의 인권수행과정의 어려움에 대한 탐색적 연구 - 부산지역 장애인거주시설종사자를 중심으로.” 『한국콘텐츠학회논문지』 제16권 3호: 771-781.
- 김용득. 2016. “지역사회중심 장애인서비스 정책의 쟁점과 과제.” 『사회서비스연구』 6권 2호: 1-28.
- 김용진·조원일. 2018. “장애인복지 패러다임 변화에 따른 장애인거주시설 사무국장의 장애인 자립관에 대한 실체와 체화 경험.” 『한국사회복지질적연구』 제12권 2호: 127-153.
- 김형식. 2008. “UN 장애인 권리협약과 장애인의 시민적 권리.” 『재활복지』 제12권 3호: 143-170.
- 남구현·박숙경·김명연·임성만·박경석·박래균·염형국·박옥숙·여준민·임소연·김정하·김주현·권미진·손현희. 2005. 『장애인생활시설 생활인 인권상황 실태조사: 양성화된 조건부신고복지시설을 중심으로』. 장애우권익문제연구소·국가인권위원회.
- 박경수·장혜경. 2013. “장애인 거주시설 실천가의 인권 딜레마 사례연구.” 『한국장애인복지학』 제22권: 249-272.
- 박선영. 2013. “사회복지실천의 임파워먼트 접근에서 온정주의의 침투와 공존: 딜레마와 대안.” 『미래사회복지연구』 제4권 1호: 107-135.

- 임성택. 2012. “장애인 생활시설에서의 인권침해 그 현황과 대책.” 『저스티스』 제9권: 5-38.
- 이선우. 2009. 『장애인복지의 이론과 실제』. 서울 : 집문당.
- 이소영. 2011. “OECD 국가의 장애정책 패러다임의 전환.” 『글로벌사회복지연구』 제6권:31-52.
- 이종정. 2018. “장애모델의 발전사를 통해 살펴 본 긍정적 장애모델의 특성과 함의.” 『한국장애인복지학』 41권: 165-194.
- 이지수. 2014. “장애의 사회적 모델이 가진 의미와 한계 : 장애 개념정의와 정체성의 정치를 중심으로.” 『한국장애인복지학』 25권: 33-54.
- 전광석. 2011. 『한국사회와 장애인 정책: 복지와 차별금지』. 서울 : 인간과 복지.
- 정종화·이종길·이경준·조석영·김상진·김선·이소영·정유진·최선자·함석홍. 2009. 『장애인복지서비스 이용자 참여 매뉴얼』. 고양 : 공동체.
- 조우연·신경애·김영이·김은주·전은경·강희숙. 2013. “지적장애인 거주시설 사회복지사의 윤리적 딜레마 경험.” 『재활복지』 제17권 1호: 233-258.
- 조한진·고영신·곽정란·김승엽·김희선·임소연·최희정. 2006. 『사례분석을 통한 지역사회 정신지체장애인 인권침해실태 연구』. 장애우권익문제연구소·국가인권위원회.
- 조효제. 2009. 『인권 딜레마』. 인권대학 ‘들숨날숨, 인권과 호흡하기’ 자료집.
- Colaizzi, P. F. 1978. “Psychological research as a phenomenologist views it.” In: Valle, R.S. and King, M., Eds., *Existential-Phenomenological Alternatives for Psychology*. Oxford University Press. New York. 48-71.
- Doyle, B. 1999. “From welfare to rights? Disability and legal change in the United Kingdom in the late 1990s.” In M. Jones and L.A.B. Marks(eds), *Disability, Divers-ability and Legal Change*. The Hague: Kluwer Law International/ Martinus Nijhoff Publishers, 209-26.
- Giorgi, A. 1985. *Phenomenology and psychological research*. Pittsburgh, Duquensne university press.

- Healy, L. M. 2008. "Exploring the history of social work as a human rights profession." *International Social Work*. 51(6): 735-748.
- Ife, J. 2001. *Human Rights and Social work*. Columbia University Press.
- Krueger, R. A. 1988. *Focus groups: A practical guide for applied research*. California: SAGE Publications, Inc.
- Lincoln, Y. S., and Guba, E. G. 1985. *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Morgan, D.L. 1996. *Focus groups as qualitative research*. Newbury Park. CA: Sage.
- Waddington and Diller. 2002. "Tensions and coherence in disability policy: The uneasy relationship between social welfare and civil rights models of disability in American, European and International Employment Law." *Disability Rights Law and Policy: International Perspectives*. 241-282.

A Qualitative Case Study on the Human Rights based Approach in Living Facility of Persons with Disabilities: Focusing on the Experiences of Social Workers*

Jihye Jeon**

Abstract

This study explored the dilemma experienced by social workers in living facilities of people with disabilities, and discussed the contexts why these dilemmas were caused. The 3 focus groups interviews of social work practitioners were conducted. The analysis revealed that the dilemma situation contains three dimensions : dimension of position and role value, dimension of philosophy, dimension of social work practice. Although the disability paradigm is shifted from welfare approach to human right approach, the social work practice in living facility of people with disabilities is incompatible with human rights approach. Therefore, this study emphasized human right education to both social workers and people with disabilities for enhancing their relationship and understanding, and the development of communication skill and practical rules in conflict situation.

Keywords: human right based approach, disability paradigm shift, living facility for people with disabilities, dilemma, social work practice

투고일: 2019.02.14. 심사일: 2019.03.15. 게재확정일: 2019.03.15.

* This work was supported by Incheon National University Research Grant in 2017.

** Assistant Professor, Incheon National University.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.93>

【연구논문】

공공기관 고객만족지수(PCSI)모형을 활용한 기록정보서비스
품질조사: P기록관을 중심으로*

박 제 우**, 윤 유 라***

논문요약

공공기관의 경영 패러다임이 이용자 중심으로 전환되면서 이용자를 대상으로 하는 서비스 품질 평가의 중요성이 날로 커지고 있다. 그러나 타 기관과 비교하여 기록관의 경우 서비스 품질에 대한 인식 및 연구가 부족하였던 것이 사실이다. 본 연구에서는 이러한 문제점에 대해 인지하고, 기록관의 서비스 품질 측정의 필요성을 논의함과 더불어 기존에 활용되고 있는 품질평가 기법을 사용하여 기록관의 정보서비스 품질을 직접 측정해보고자 하였다. 이를 위해 사용한 품질평가 방법이 PCSI다. PCSI는 공공기관에서 서비스 만족도를 측정하고 만족도를 향상시킬 수 있는 방법을 찾고자 고안된 모델이다. 본 연구에서는 PCSI에 대해 분석하고 기록관에 적용할 수 있도록 평가지표를 수정하였다. 이를 바탕으로 P기록관을 대상으로 설문조사를 실시하여 이용자 중심의 기록정보서비스 품질 평가를 진행하였다. 조사결과 먼저 만족 모형을 통하여 산출된 P기록관의 고객만족도 점수는 80.2점으로 나타났으며, 만족 지수 중 이용자들은 환경품질에 대해 가장 큰 만족을 느끼는 것으로 나타났다. 둘째, 서비스 품질에 대해 조사한 결과 상품품질에 대한 평가에서는 ‘전문성’이 최고점을 받아 강점으로 파악되었으며, 최저 점수를 받은 혁신성이 약점으로 파악되었다. 전달품질에서는 담당자의 공감성이 최고점을 대응성이 상대적으로 최저점을 받았다. 환경품질 영역에서는 쾌적성이 최고점을 편리성이 최저점으로 나타났다. 사회품질에서는 직원의 태도가 가장 높은 점수를 받았으며, 안전성은 가장 낮은 점수로 나타났다. 본 연구 결과는 이용자 기반의 기록정보서비스를 개발하고 만족도를 향상하기 위한 기초 데이터로 활용될 수 있을 것이라 기대된다.

주제어: 공공기관 고객만족지수, 기록정보서비스 품질평가, 기록관

* 이 연구는 동의대학교 대학원 석사학위논문의 일부를 요약수정한 것임.

** 동의대학교 문헌정보·사학과 기록관리전공 석사.

*** 동의대학교 문헌정보학과 조교수, 교신저자.

I. 서론

오늘날 공공기관은 기존의 기관 중심이 아닌 이용자 중심의 경영 패러다임으로 전환되고 있으며, 품질평가 또한 기관의 성과에 중점을 두었던 과거와는 달리 이용자 중심의 서비스 품질평가로 변화되고 있다. 그러나 이러한 공공기관의 변화에도 불구하고 기록관의 변화는 상대적으로 적은 것이 사실이다.

국가기록원의 2017년 공공기관 기록관리 평가결과를 살펴보면(국가기록원, 2017), 기록정보서비스 및 업무개선 분야보다는 여전히 기반구축 및 보존에 집중하고 있음을 알 수 있었다. 서비스 및 업무개선 분야 점수는 전체 평균인 73.8점보다 한참이나 모자란 점수인 60.1점을 받았으며, 이는 업무기반(77.5점) 및 업무추진 분야(74.9점) 점수와도 큰 차이가 있는 것으로 나타났다.

기록관의 궁극적인 목적이 단순히 기록의 보존에 국한되지 않고, 서비스를 받는 이용자들의 이용에 있음을 생각한다면(설문원, 2008) 향후, 기록관에서도 서비스 개발과 제공에 관한 계획을 갖추고 서비스 품질을 향상시킬 방법에 관해서 치밀하게 고민할 필요가 있다. 이를 위해서는 이용자 중심의 사고를 통해 기관이 가진 내·외부의 강점 및 약점에 대해 평가하고 기관의 전략적 계획이 세워져야 한다. 즉, 이용자 중심의 기관평가가 도입되고 이를 바탕으로 한 서비스 개선이 이루어져야 한다.

본 연구에서는 향후 기록관이 제공하는 기록정보서비스 품질 향상을 위한 필요성을 인식하고, 기록정보서비스의 특성을 반영한 평가지표를 활용하여 이용자중심의 기관 평가를 실시해보고 기록정보서비스 품질의 실태를 파악하는 것에 목표를 두고자 한다.

현재 국내 공공기관에서는 공공기관 고객만족지수(PCSI : Public-service Consumer Satisfaction Index, 이하 PCSI)를 활용하여 기관을 평가개선하고 있다. 이에 본 연구에서도 PCSI를 활용하여 실제 단위 기관인 부산의 P기록관의 기록 정보 서비스에 대해 평가해 보고 고객의 만족과 불만족 요소를 분석해보고 향후 개선방안에 대해 살펴보려고 한다.

II. 공공기관 고객만족지수 모형의 개념과 사례

1. 공공기관 고객만족지수(PCSI)모형

공공기관에서는 효율적 운영과 생산성을 높이기 위한 평가의 중요성이 높아짐에 따라 민간부문의 국가고객만족지수인 NCSI(National Customer Satisfaction Index, 이하 NCSD)를 활용하여 공공기관의 고객만족도를 측정하고자 하였다.

그러나 NCSI와 기존의 고객만족도지수모형인 KCSI(Korea Customer Satisfaction Index, 이하 KCSI)는 민간부문을 주 대상으로 하는 모델이기 때문에 공공기관의 공공성이라는 특수성을 반영하기 어렵다는 한계점이 있었다. 예를 들면, <표1>처럼 민간부문과 공공부문은 기본적으로 고객에 대한 정의와 가치기준이 상이하며, 나아가 민간의 경우 고객의 재화 및 서비스 재구매 행위를 성과로 평가하는 반면, 공공기관은 고객의 행복과 사회적 기여를 성과의 결과로 보고 있다는 점에서도 차이를 보이고 있다.

<표 1> 민간부문과 공공부문 조직 목표의 차이

	민간부문	공공부문
고객	고객(Client)	시민(Citizen) 또는 국민
가치	Private Value	Public Value
서비스 품질 결과	Customer Satisfaction	People satisfaction
사회교환	상품, 재화, 금전 등의 유형 자산	협조, 존경, 타협
고객만족의 결과변수	재구매, 구전	국민행복, 사회적 기여

*출처 : “공공기관 고객만족지수 모형의 개발과 적용: PCSI (Public-service Customer Satisfaction Index)를 중심으로” 이유재 외, 2012

이에 따라 2004년 한국능률협회컨설팅(KMAC)과 서울대학교가 공동연구를 진행하여, 기존의 고객만족지수 모형인 NCSI, KCSI, 한국표준협회의 KS-SQI모형

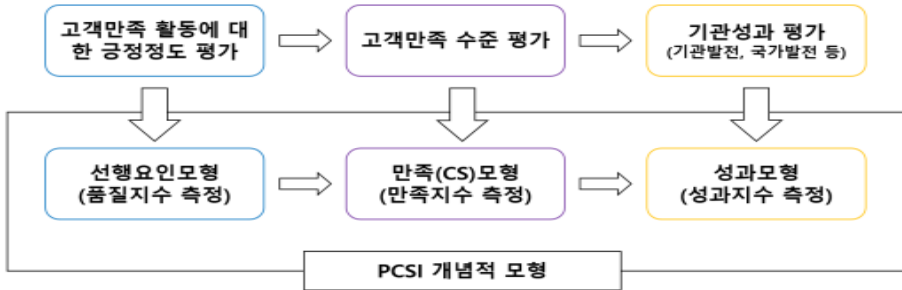
(Korean Standard-Service Quality Index), 서울시민 만족도 조사인 SSI(Seoul Satisfaction Index) 등을 분석하여 공공기관이 가지는 특수성을 반영한 공공기관 고객만족지수(Public-service Customer satisfaction Index, 이하 PCSI)모형을 개발하였다. 이후 공공기관은 국민을 대상으로 연 1회 이상의 고객만족도 조사를 실시해야 한다는 「공공기관의 운영에 관한 법률」이 개정됨에 따라¹⁾ 초기 모형을 개편한 PCSI 모형이 공공기관 고객만족도 평가를 위한 모델로 채택되어 현재까지 활용되고 있다. (이유재 외, 2012)

2. PCSI의 개념 및 구조

PCSI모형은 이용자의 만족이 서비스의 품질에서 비롯되며 이러한 만족은 공공기관의 성과로 이어진다는 인과관계 개념을 적극적으로 도입한 평가모델이다. 이러한 인과관계는 <그림 1>과 같이 표현될 수 있다. 공공기관의 품질 즉 선행요인에 의해 고객은 만족·불만족을 느끼게 되며 이는 실질적인 기관의 성과결과에 영향을 미치게 된다는 의미이다. 즉, 만족점수가 도출된 요인인 서비스 품질지수(선행요인) 분석을 통해 자관의 문제점이 무엇인지 보다 구체적으로 파악할 수 있으며, 이에 대한 개선을 통해 자연적으로 만족지수 및 성과지수가 상승될 수 있도록 만들어진 유기성을 확보한 모델이다. 이처럼 PCSI모형은 단순히 만족지수를 평가하기 위한 모델이라기보다는 각 기관의 서비스 품질의 강점요소와 약점요소를 파악하여 개선을 유도하는 현장 개선형 모델이라 할 수 있겠다.

1) 공공기관의 운영에 관한 법률 제 13조 2항, 국민에게 직접서비스를 제공하는 공공기관은 그 공공기관의 서비스를 제공받는 국민을 대상으로 연 1회 이상 고객만족도 조사를 실시하여야 한다. 이 경우 기획예산처장관은 공공기관으로 하여금 고객만족도 조사를 통합하여 실시하게 하고, 그 결과를 종합하여 공표할 수 있다. <제정 2007.1.19.>

〈그림 1〉 PCSI모형의 인과모형 방식



*출처 : “공공기관 고객만족지수 모형의 개발과 적용: PCSI (Public-service Customer Satisfaction Index)를 중심으로” 이유재 외, 2012

이러한 PCSI모형의 장점은 기획재정부 주관으로 2017년도에 실시한 공공기관의 고객만족도 조사 결과보고서²⁾에서도 나타난다. 보고서를 살펴보면 만족도점수가 점차적으로 상승되고 있음을 알 수 있다(〈표2〉 참조). 물론 다양한 원인이 있겠으나, 공공기관에서 PCSI모형을 활용하여 기관에서 제공하고 있는 서비스의 개선을 꾀하였고 이러한 결과로 고객만족도 역시 상승되었다고 평가하고 있다.³⁾

〈표 2〉 기획재정부 주관 공공기관의 PCSI점수 변화

	2009	2010	2011	2012	2013	2014	2015	2016
공기업	92.0	92.9	93.7	93.0	93.9	94.4	92.4	92.5
준정부기관	86.7	88.4	89.6	89.4	89.7	89.9	86.6	87.1
기타공공기관	82.7	84.7	86.9	85.1	86.3	86.3	85.8	85.1

*출처 : 〈공공기관 고객만족도 결과〉 기획재정부, 2018

또한, PCSI모형은 객관적이고 합의된 기준을 통해 공공기관의 평가를 진행하기 때문에 타 기관과 자관의 서비스 만족도의 차이 및 품질과 성과에 대한 점수도 비교분

2) 기획재정부 경영혁신과. 2018. 공공기관 고객만족도 결과. http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1103(검색일: 2018.12.18.).
 3) 만족지수를 높이는 방법을 파악하고 역으로 점수를 올리기 위한 방안의 하나로 PCSI를 악용하고 있다는 비판적 견해도 있다.

석 가능하다.

PCSI점수는 만족 모형의 하위요소인 전반적 만족, 요소 만족, 사회적 만족 3개의 차원에서 측정되며 구체적인 방법은 다음과 같다.

첫째, 만족 모형의 차원을 구성하고 있는 각 구성요소의 측정항목을 합산하여 평균하는 산술평균값으로 점수를 산출한다. 둘째, 구성요소의 산출된 점수를 통하여 같은 방법으로 차원별 만족도 점수를 산출한다. 셋째, 도출된 차원별 만족도 점수에 가중치를 산정하여 평균값을 계산하여 도출한다. 각 가중치는 전반적 만족 50%, 요소 만족 30%, 사회적 만족 20%이다. PCSI 산출방법은 다음과 같이 <그림 2>로 나타낼 수 있다.

<그림 2> PCSI 지수 산출방법



*출처 : “공공기관 고객만족지수 모형의 개발과 적용: PCSI (Public-service Customer Satisfaction Index)를 중심으로” 이유재 외, 2012

3. PCSI의 적용사례

지금까지 기록관리 분야에서 PCSI모형을 활용하여 서비스 품질평가를 진행한 연구는 없었다. 그러나 유관분야에서는 PCSI를 적용한 평가들이 이루어져 왔다. 유관분야 중 PCSI모형의 활용사례로는 한국과학기술정보연구원(KISTI)의 오픈엑세스 서비스인 NOS 서비스, 한국환경산업기술원, 국회도서관을 사례로 들 수 있다. 이를 정리하면 <표3>과 같다.

〈표 3〉 유관기관의 PCSI 적용사례

	현미환(2013)	고객만족도 조사 결과보고서(2013)	권나현(2018)
기관	한국과학기술정보연구원 (NOS 서비스)	한국환경산업기술원	국회도서관
평가방법	설문조사	설문조사 및 일대일 면접	설문조사
평가대상	NOS 서비스 이용자	해당 기관 서비스 이 용 경험 고객	국회도서관 서비스 이 용 경험 고객
선행요인 모형관련 항목구성	- PCSI 선행요인모형 11개요소 포함 온라인정보서비스의 특성으로 심미성 제외	- PCSI 선행요인모형 12개요소 모두 포 함.	- PCSI 선행요인모형 요소 12개요소 모두 포함.

현미환(현미환 외, 2013)은 PCSI를 통해 NOS서비스의 고객만족도를 조사하고 3가지 핵심 문제를 발견하였고 그 개선 방안으로 인증기 발급 절차를 간소화, 자동인증방식으로서의 변경, 검색식 작성 및 활용법 안내를 통해 서비스 품질을 개선하고 고객만족도를 높일 수 있을 것으로 분석하였다.

한국환경산업기술원은 전달품질 중 가장 점수가 낮았던 대응성을 높이기 위한 업무별 핵심 개선 과제를 선정하고 하드웨어적, 소프트웨어적, 휴먼웨어적 중요 개선 필요 부문을 서비스 품질 개선 방향으로 제시하였다.

국회도서관을 대상으로 품질 평가를 실시한 권나현(2018)은 혁신성이 가장 낮음을 밝혀내고 국회도서관 전체 서비스 품질과 부서별 서비스 품질의 개선 필요성을 제시하였다.

이와 같이 PCSI모형은 공공기관들이 동일한 모형을 적용해 결과에 대한 비교가 가능하며, 서비스 품질의 강점과 약점을 분석할 수 있어 서비스 품질의 개선을 위한 방안을 제시할 수 있다는 장점이 있다.

PCSI모형은 새로운 환경에 맞추어 개선점을 찾고 변화하고 있다. 이에 현재 자관의 상황에 맞추어 세부 평가요소의 비율을 조정할 수 있도록 개선한 PCSI 2.0 버전이 개발되었다. 그러나 기관의 내부자가 아닌 이상 기관별 세부적 중점분야를 선택하는 것에는 어려운 점이 있어 본 연구에서는 통일된 분석요소를 제안하고 있는 PCSI 1.0버전을 활용하였다.

Ⅲ. 연구설계

1. 조사기관 개요

본 연구에서는 기록관 중 영남지방의 대표적인 기록관인 P기록관을 대상으로 연구를 진행하였다. P기록관은 1984년 건립 후 2008년 리모델링을 통해 현재까지 영남권의 국가기관, 지방자치단체, 공공기관 및 민간에서 생성한 중요 기록물을 수집, 보존, 활용하고 있다. 기록관 내의 시설로는 서고 28개를 운영 중에 있으며, 기록문화전시관 2개(기록문화역사실, 조선왕조실록실)와 기록정보열람실이 있다.

P기록관의 기록정보서비스는 온/오프라인 열람서비스를 비롯하여 전시, 체험교실, 견학 등 다양하다. 본 연구에서는 선행연구 분석을 바탕으로 기록정보서비스를 온/오프라인 기록열람서비스, 복사서비스, 기록관 안내, 이용자 면담, 검색도구 제공 등을 통하여 기록관이 소장하고 있는 기록물에 대한 정보를 제공하고 접근 활동을 지원하는 활동을 참조서비스로, 온/오프라인 전시, 교육, 방문자에 대한 안내, OT, 교육 등 기록물의 가공, 해석 작업을 통해 이용자들이 기록물을 쉽게 접할 수 있고 의문을 해결할 수 있도록 지원하는 활동을 부가가치서비스로 문화행사를 통한 기록관에 대한 홍보, 기록과 기록관리에 대한 홍보, 매체 홍보, 견학, 마케팅, 서포터즈, 자원봉사자 출판 등을 포함하여 기록관이 기록관과 기록관리의 존재와 중요성을 알리는 활동 등을 확장서비스로 범주화하여 구분하였다.(Pugh, 2005; 이숙희, 2004; 설문원, 2008; 이명희 외, 2008; 허덕행, 2014)

이러한 구분을 바탕으로 P기록관의 주요 기록정보서비스를 범주화하면 <표 4>와 같다.

<표 4> P기록관의 기록정보서비스 유형

유형	P기록관의 기록정보서비스
참조서비스	온/오프라인 열람서비스, 스캐닝, 복사, 소장기록물 안내, 이용안내 등
부가가치서비스	전시, 기록문화 체험교실 등
확장서비스	견학, 팸플렛, 리플렛 등 참조자료 발간, 기록사랑 전국 백일장, 진로체험교실, 기록교실 등

2. 조사문항

설문조사를 위한 평가지표의 구성은 <표 5>와 같다.

첫 번째로 기관이 제공하고 있는 서비스의 품질을 측정하는 선행요인모형은 상품품질, 전달품질, 환경품질, 사회품질 4개 차원으로 구성되어 있으며, 각각의 차원은 3개의 하위요소로 구성되어 있다. 이에 본 연구에서는 총 22개의 문항으로 구성하였다. 두 번째로 고객 또는 이용자의 만족 정도를 평가하기 위해 전반적 만족, 요소 만족, 사회적 만족의 3개 차원과 8개의 하위 항목으로 구성되어 있는 만족모형은 총 7개의 문항으로 구성하였다. 마지막으로, 선행요인과 만족 요인을 기반으로 기관성과와 공공기관의 서비스가 사회발전에 기여하는 영향을 측정하는 성과모형은 4개의 문항으로 구성하였다. 이외에 응답자의 기본사항을 묻는 질문 6개를 포함하여 총 39개의 문항으로 평가지표를 구성하였다.

각 항목은 Likert 5점 척도를 사용하여 매우 만족(5점)부터 매우 불만족(1점)으로 평가하도록 하였다.

<표 5> PCSI를 적용한 설문 구성

모형	차원	구성요소	측정항목	문항수
서비스 품질 모형	상품품질 (서비스 내용)	편의성	1. 혜택과 가치 제공 유·무 2. 서비스 내용의 유용성 유·무 3. 시간과 비용의 절감성 유·무 4. 정보제공의 적합성 유·무	4개
		전문성	1. 담당자의 업무에 대한 지식, 기술, 전문성 보유 유·무 2. 정보제공을 받기위한 전문적인 시스템 보유 유·무	2개
		혁신성	1. 이용자를 위해 새로운 아이디어나 프로그램 기획 유·무	1개
	전달품질 (서비스 전달)	대응성	1. 처리절차 또는 구비서류의 간편성, 신속성 유·무 2. 처리 결과 통보의 신속성 유·무	2개
		지원성	1. 처리 후 확인 점검 혹은 불만 사항에 대한 대응성 유·무	1개

		공감성	1. 이용자에 대한 담당자의 친절성 유·무	1개
	환경품질 (서비스 환경)	쾌적성	1. 기록관 환경의 쾌적성 유·무 2. 기록관 환경의 관리성 유·무	2개
		편리성	1. 기록관 시설의 편리성 유·무 2. 이용 및 서비스를 제공받기 위한 창구의 편리성 유·무 3. 서비스 담당 부서나 업무 담당자와의 연락 용이성 유·무	3개
		심미성	1. 홈페이지의 심미성 유·무 2. 시설의 안락성 유·무	2개
		사회품질	직원의 태도	1. 직원들의 공정성 유·무
	공익성		1. 기관의 지역 사회의 발전에 기여도 유·무	1개
	안전성		1. 정보 보호의 신뢰성 유·무 2. 안전을 위한 노력성 유·무	2개
서비스 만족 모형	전반적 만족	절대적 만족	1. 기록관 서비스의 전반적 만족도 정도	1개
		상대적 만족	1. 서비스를 제공받기위한 투자 대비 만족도 정도	1개
		감정적 만족	1. 서비스를 이용하신 후 감정적 만족도 정도	1개
	요소 만족	상품만족	1. 서비스 내용에 대한 만족도 정도	1개
		전달만족	1. 서비스 전달과정에 대한 만족도 정도	1개
		환경만족	1. 서비스 환경에 대한 만족도 정도	1개
	사회적 만족	사회적 책임만족	1. 사회적 책임과 역할 만족도 정도	1개
성과 모형	기관성과	신뢰도	1. 서비스 이용 후 P기록관의 신뢰성 정도	1개
		호감도	1. 서비스 이용 후 재이용 의향 유·무	1개
		구전공정 도	1. 서비스 이용 후 타인에게 추천 의향 유·무	1개
	사회성과	국가발전	1. P기록관이 국가의 발전기여도 유·무	1개

3. 조사방법

PCSI를 활용한 P기록관의 기록정보서비스품질 조사는 기록관을 방문한 이용자를 대상으로 직접설문 및 간략 면담형식으로 진행하였다. 조사기간은 2018년 11월 03일에서 17일까지 총 2주간 실시하였다. 설문지는 열람서비스와 부가가치서비스 등 기록관 서비스를 이용하고 나온 이용자를 대상으로 하여 직접 배포, 수거하였고 이 중 일부 이용자를 대상으로 간략 인터뷰를 진행하였다. 수집된 데이터는 SPSS 25 통계 분석 프로그램을 사용하여 분석하였다.

IV. 조사 결과

1. 응답자 특성

응답자의 인구통계학적 사항으로는 성별, 나이, 직업을 조사하였고 결과는 다음과 같다. 총 75명 응답자의 성별 분포는 여성이 37명(49.3%), 남성이 38명(50.7%)의 분포를 보였다. 연령별 분포는 20대가 27명(36%)으로 가장 많았고 30대가 16명(21.3%), 40대 15명(20%), 50대 7명(9.3%), 10대 6명(8%), 60대 이상이 4명(5.3%)로 나타났다. 응답자의 직업은 대학생/대학원생이 23명(30.7%)로 가장 많았고, 회사원 13명(17.3%), 공무원 11명(14.7%), 기타 11명(14.7%), 주부 12명(9%), 중학생/고등학생 6명(8%), 교수 및 연구자 2명(2.7%)의 순으로 나타났으며 요약하면 <표 6>와 같다.

〈표 6〉 응답자의 인구통계학적 특성

특성	세부항목	빈도(n)	비율(%)
성별	여성	37	49.3
	남성	38	50.7
연령	10대	6	8.0
	20대	27	36.0
	30대	16	21.3
	40대	15	20.0
	50대	7	9.3
	60대 이상	4	5.3
직업	중학생/고등학생	6	8.0
	대학생/대학원생	23	30.7
	교수 및 연구자	2	2.7
	주부	9	12.0
	회사원	13	17.3
	공무원	11	14.7
	기타	11	14.7
합계		75	100.0

응답자들이 이용한 서비스가 무엇인지 알아보았으며 그 결과는 〈표 7〉과 같다. 응답자 중 확장서비스 이용자는 51명(68.0%), 열람서비스 이용자는 24명(32.0%)으로 나타났다.

〈표 7〉 응답자의 이용 서비스

특성	세부항목	빈도(n)	비율(%)
이용 서비스	열람서비스	24	32.0
	부가가치서비스	0	0.0
	확장서비스	51	68.0
	기타	0	0.0
합계		75	100.0

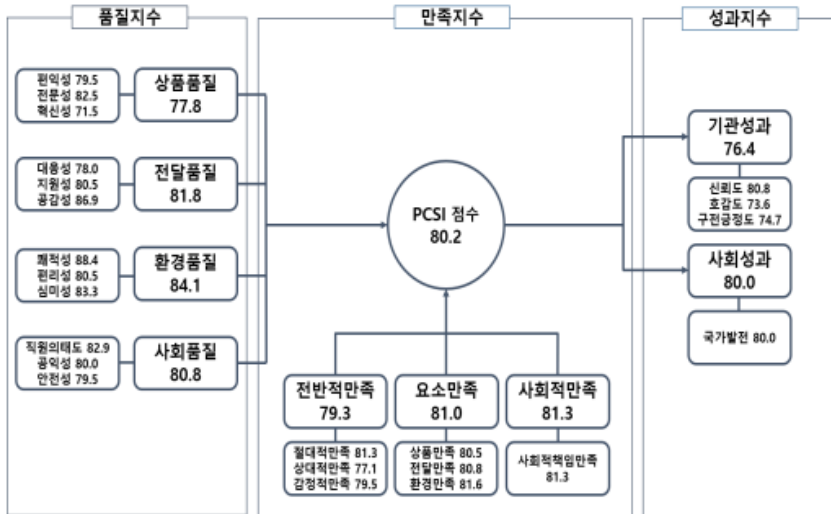
2. 종합적 분석

PCSI 조사 결과는 <그림 3>과 같다. P기록관 이용자의 고객만족도 점수는 80.2점으로 나타났다. 본 연구의 목적에 맞춰 만족지수의 요소 만족 차원과 품질지수를 중점으로 분석하였다.

각 지수별 분석결과는 다음과 같다.

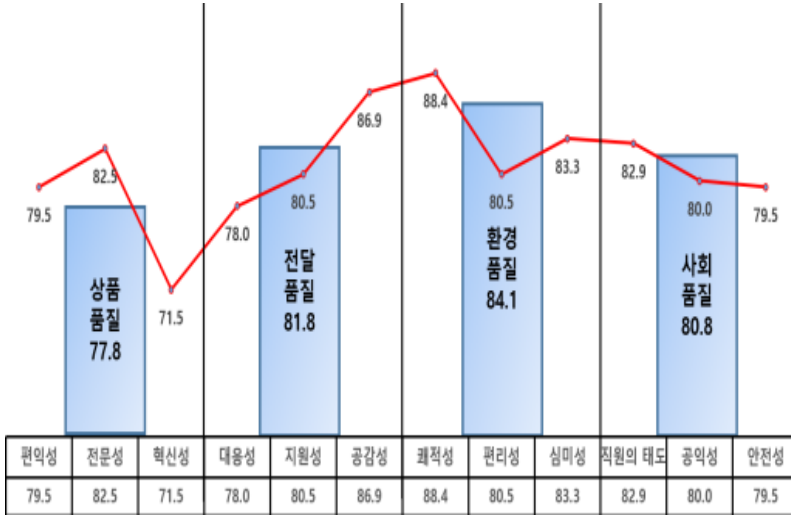
첫째, 만족지수에서는 사회적 만족이 81.3점으로 가장 높게 나타났으며, 요소 만족이 81.0점, 전반적 만족이 79.3점으로 가장 낮게 나타났다. 요소 만족에서 이용자들은 서비스 환경품질에 가장 큰 만족도(81.6점)를 느끼고 있었으며, 상품품질에 대한 만족도(80.5점)가 가장 낮은 만족도를 보인 것으로 나타났다.

<그림 3> P기록관 이용자의 PCSI 결과



둘째, 품질지수에 대한 구체적인 조사결과는 <그림 4>와 같다. 품질지수에서는 환경품질 84.1점으로 가장 높은 점수로 나타났고, 전달품질 81.8점, 사회품질 80.8점, 상품품질이 77.8점의 순으로 나타났다.

〈그림 4〉 서비스 품질 차원의 평균점수와 구성요소 점수



품질지수에서는 상품품질의 담당자와 시스템의 전문성에 가장 높은 점수(82.5점)로 나타났으며 새로운 아이디어나 프로그램의 기획에 대한 혁신성에 대해서는 가장 낮은 점수(71.5점)로 나타났다.

혁신성에 관하여 부정적 응답으로 대답한 일부 응답자를 대상으로 간략 인터뷰를 진행하였으며, 기관의 홍보방식으로 인해 불편함을 겪고 있는 것으로 나타났다. 관련 기타의견은 간략 인터뷰를 통해 다음과 같이 나타났다.

“매년 학생들의 진로를 위해 기록관을 방문합니다. 하지만 견학 내용이 바뀐게 있는지, 새로운 프로그램은 없는지 매년 사전답사를 해야하는 불편함이 있어요. 제 성격이 꼭 확인해야 되는 성격이긴 하지만 그렇게 하지 않으면 변한게 있는지 딱히 확인할 방법이 없거든요. 대부분 그래서 동료들에게 확인하거나 방문한 뒤 신청을 하는 것 같아요..”(이용자 A, 2018년 11월7일)

전달품질에서는 담당자의 친절성을 측정하는 공감성(86.9점)에 대해 높은 점수를 주었으나 구비서류의 불편함과 처리 결과 통보의 신속성을 측정하는 대응성에 낮은 점수(78.0점)를 주었다.

대응성에 관하여 부정적 응답으로 대답한 일부 응답자를 대상으로 간략 인터뷰를 진행하였으며, 절차와 구비서류의 복잡함, 수수료 발생으로 인해 불편함을 겪고 있는 것으로 나타났다. 관련 인터뷰 내용은 다음과 같다.

“저도 기록관의 존재는 다른 기관을 방문하여 알게 되었어요. 나와 직접적으로 관련된 서류를 찾는 데도, 조상님의 토지문서를 찾으려고 하는 건데도 필요한 서류와 절차는 너무 복잡했어요.”(이용자 B, 2018년 11월 16일)

“기록관 방문이 이번에 벌써 3번째예요 3번째 그런데 일부러 직접 기록관에 방문해서 기록물을 보려고 하는 편입니다. 온라인 신청이나 FAX로 받으려는 경우 수수료가 생각보다 만만치 않더군요. 근데 제일 이해하기 어려웠던 것은 기록물을 확인하는데 절차가 너무 많고 불편하더군요.”(이용자 C, 2018년 11월 16일)

환경품질에서는 기록관 환경이 쾌적하고 환경의 관리성에 높은 점수(88.4점)를 준 반면에 담당자와의 연락 용이성 등을 측정하는 편리성에는 낮은 점수(76.8점)를 받았다.

사회품질에서는 직원들의 공정성을 측정하는 직원의 태도가 가장 높은 점수(82.9점)로 나타났으며 안전을 위한 노력 등을 측정하는 안전성(79.5점)은 가장 낮은 점수로 나타났다.

셋째, 성과지수에서는 기관성과 76.4점, 사회성과 80.0점으로 나타났다. 기관성과에서는 신뢰도가 80.8점으로 가장 높게 나타났으며, 호감도가 73.6점으로 가장 낮은 점수로 나타났다.

V. 결론

본 연구에서는 우리나라 공공기관에서 사용하고 있는 PCSI모형을 활용해 기록관리 분야에 적용해보고자 하였다. PCSI모형은 품질모형, 만족지수 모형, 성과지수 모형으로 구성되어 있으며, 고객 만족의 선행요인으로 서비스 품질모형을 두고, 만족 모형과

서비스 품질모형의 결과요인으로 성과모형을 두는 인과구조를 갖추고 있어 서비스 품질 개선의 분석이 용이하다는 강점을 가지고 있다. 이를 위해 부산지역의 P기록관을 대상으로 설문조사를 실시하여 서비스 품질의 강점과 약점을 파악하고 서비스 품질 개선의 요소가 무엇인지 분석하였다.

조사결과 P기록관의 PCSI점수는 80.2점이었다. 2017년 기획재정부 주관의 공공기관의 고객만족도 결과(기획재정부, 2017)를 보면 공공기관의 고객만족도 평균점수는 공기업 92.5점, 준정부기관 87.1점, 기타공공기관 85.1점으로 나타났다. 본 연구의 조사결과와 비교하였을 때 P기록관의 고객만족도 평균점수는 타 공공기관에 비하여 비교적 낮은 것으로 나타났다.

향후, P기록관의 고객만족도를 체계적으로 향상시키기 위한 서비스 개선점을 파악하기 위해 서비스 품질의 강점과 약점을 정리하면 다음과 같다. 상품품질에서는 전문성(82.5점)이 강점으로, 혁신성(71.5점)은 약점으로 확인되었고, 전달품질에서는 공감성(86.9점)이 강점으로, 대응성(78.0점)은 약점으로 확인되었다. 환경품질에서는 쾌적성(88.4점)이 강점으로, 편리성(80.5점)이 약점으로 확인되었고 사회품질에서는 직원의 태도(82.9점)는 강점으로 안전성(79.5점)이 약점으로 측정되었다.

이러한 결과를 통해 향후 P기록관의 고객만족도를 향상시키기 위해서는 새로운 서비스에 대한 개발과 더불어 서비스에 대한 홍보활동을 강화는 물론, 서비스를 이용하는 과정의 단순화 및 신속성을 강화하기 위한 방안이 마련되어야 할 것으로 판단된다.

참고문헌

- 국가기록원. 2017. 『2017년도(‘16년 실적) 기록관리 평가결과 보고.』
- 권나현·이정연·표순희. 2018. “공공기관 고객만족지수(PCSI) 모형을 적용한 도서관 고객만족도 측정.” 『한국문헌정보학회지』 제52권 제1호: 313-340.
- 설문원. 2008. “기록정보서비스 전략계획 수립을 위한 기초 연구:국가기록원을 중심으로.” 『정보관리학회지』 제25권 제3호: 249-271.
- 설문원. 2008. “기록정보서비스의 방향과 과제.” 『기록인』 (여름): 10-19.
- 이숙희. 2004. “기록물관리기관의 열람환경 운영에 관한 연구.” 『기록학연구』 제10호: 65-111.
- 이유재·이청림. 2012. “공공기관 고객만족지수 모형의 개발과 적용: PCSI (Public-service Customer Satisfaction Index)를 중심으로.” 『마케팅연구』 제27권 제4호: 69-99.
- 장혜란·이명희. 2008. “지방기록물관리기관을 위한 이용자 서비스 모형 개발에 관한 연구.” 『정보관리학회지』 제25권 제3호: 321-338.
- 허덕행. 2014. “기록정보서비스 평가 지표에 관한 연구.” 명지대학교 석사 학위논문
- 현미환·김완중·이혜진·김혜선. 2013. “PCSI를 이용한 과학기술정보 오픈서비스 이용자 만족도에 관한 연구.” 『한국정보관리학회지』 제30권 제4호: 133-154.
- KMAC. 2013. 『준정부기관 고객만족도 조사 결과보고서.』
- Pugh Mary Jo 저. 설문원 역. 2004. “Providing Reference Services for Archives and Manuscripts.” 서울: 진리탐구.
- 기획재정부 경영혁신과. 2018. 공공기관 고객만족도 결과. http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1103(검색일: 2018.12.18.).

Measuring of Recording Information Service Quality Using the PCSI(Public Customer Satisfaction Index) : Case of the P Archives

Je-woo Park*·You-Ra Youn**

ABSTRACT

Upon the user-centered shift in management paradigm of the public institutions, the importance of service quality evaluation is emphasized over time.

Addressing such issue, this study aims to directly measure the Information and service quality of the archives using the previously established quality evaluation method. In addition, this study discusses the need for the evaluation of the service quality of the archives. The quality evaluation method used is called PCSI (Public Customer Satisfaction Index), which is a model developed for evaluating service satisfactory level and its enhancement in public institutions. In this study, PCSI was first analyzed and indicators of evaluation were modified in order to be applied to the archive. Based on this, a survey was made on the P archives to evaluate the quality of the user-centered recording Information service. The survey yielded four main results. First, the customer satisfaction score was calculated to be 80.2 from the satisfactory model, and customers showed highest satisfaction in the environment quality. In the service quality aspect, “professionalism” and “innovation” were shown to be the strength and weakness of the product quality, respectively. In the quality of delivery, the empathy of the person in charge received the highest and the correspondence received the lowest score. In the environmental quality, “comfort” received the highest score while “convenience” received the lowest score. In social quality, the records manager's sense of duty and safety each received a highest and lowest score. This study is expected to be used as a basic data for the development of user-based assement of recording Information service and enhancement of user satisfaction.

Keywords : Public Customer Satisfaction Index, Recording Information Service Quality Measuring, Archives

투고일: 2019.02.14. 심사일: 2019.03.09. 게재확정일: 2019.03.12.

* Records Management Master, Dong-eui University.

** Assistant professor, Dong-eui University.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.111>

【연구경향】

An Introduction to Statistical Causal Inference: Concepts, Models and Some Methodological Issues*

Byung-Jae Lee**

ABSTRACT

Causal inference is the most important aim in social science. Over the last 30 years, we have witnessed a plenty of conceptual and methodological improvements to deal with causal inference in statistics, social and biomedical sciences. This paper aims to provide an overview of causal modeling, which includes the counterfactual conception of causality, potential outcomes framework, and statistical models commonly used in social and biomedical sciences. Additionally, some recent issues regarding mediation analysis and structural equation models are discussed.

Keywords: Causal Inference, Propensity Score Matching, Marginal Structural Model, Causal Mediation

I. Introduction

Suppose a (overly) simplistic scenario in which a country maintains a stable democracy(or economic development) after a constitutional change of

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2016S1A3A2925033).

** Research Professor, Center for Digital Social Science, Yonsei University.

electoral system(Lindberg 2009; Persson and Tabellini 2005). Can we say that the constitutional change caused the stability(or economic development) in this country? A natural way to think would be to ask what would have happened to the country had the elections not been held. If the country would have maintained democratic stability(economic development) anyway, we would not say that the constitutional change was the cause of democracy or economic development. If, on the other hand, the country would not have maintained the stability or economic development without the constitutional reform, then we would say that the constitutional change caused the stability. Here, stability(or economic development) is the outcome, and the constitutional change is the treatment or action. To determine whether a treatment or an action causes an outcome, we typically make a mental comparison between the two scenarios; one in which the treatment is present and the other in which the treatment is absent. If the outcome differs between the two scenarios, we say that the treatment has a causal effect on the outcome. The potential outcomes framework formalizes this intuition of causality, and in some sense, one can say that the framework is similar to the most similar systems design(or methods of difference) in comparative politics(Mill 1888; Sekhon 2008).

Assessing causal effect is the primary purpose of social science across disciplines. Although many statistical models - notably linear regression models - strive to provide causal relationships among variables of interest, few can successfully offer estimates with a causal connotation. One of the main reasons for such difficulties lies in *confounding*, either observed or otherwise. Unless such confounding factors are all identified and/or controlled for, the observed association cannot be attributed to causation.

One of the commonly used approaches to address such bias is to control for covariates in the analysis(Morgan and Winship 2014). Ideally speaking, if

one can find *all* confounders for the relationship of interest, the simple calculation of the difference between treatment and control groups represents causal effects. However, as collected data and our understanding on relationships between the variables are generally limited, some bias inevitably remains due to exclusions of some important confounding variables in the analysis. Without being able to assess the effect of such bias, it would still be difficult to interpret findings from such conventional methods. A solid concept of causation is needed to assess the bias.

Randomized Controlled Trials(RCTs) are known to provide golden standard to causal inference. Although observational studies are more prone to selection bias, RCTs are not completely immune to confounders. The primary sources of confounders for RCTs are treatment *noncompliance*(not following the assigned treatment) and *missing follow-ups*(attrition from the sample). Although longitudinal models can effectively address the issue of attrition, the intention-to-treat(ITT) approach based on the treatment assigned rather than eventually received generally fails to deal with the noncompliance issue(Hollis and Campbell 1999)

Another interesting problem of importance for both experimental and observational studies is the causal mechanism of treatment effect. Although the ITT and other methods provide useful tools for estimating treatment effect, they do not tell us anything about causal mechanisms. One mechanism of particular interest is *mediation*, a process that describes the pathway from the intervention to the outcome of interest. Causal mediation analysis allows one to ascertain causation for changes of implicated outcomes along such a pathway.

The purpose of this paper is to provide a general overview of causal inference from potential outcomes perspective and to discuss some difficulty issues in causal inference. Given the expansive literature in this field, the

discussions in the following sections are inevitably very selective.

II. Counterfactual Conception of Causality and Potential Outcomes Framework

Although conceptually straightforward, a formal definition of causation is actually not (Brady 2008; Sekhon 2008). This is because we often rely on randomization for the notion of causation. How can we define causation in the absence of randomization? Since randomization is only the means by which to control for confounding, we cannot use it to define causal effect. Rather, we need a more fundamental concept to help explain why randomization can address confounding to achieve causation. This is the role of *potential outcome*.

1. Potential Outcomes and Randomized Controlled Trials

The potential outcomes framework for causal inference is based on a specific conception of causality, called *counterfactual conception of causality* (Lewis 1974; Rubin 2006; Brady 2008; Sekhon 2008). Most of recent literature on causal inference rely on the notion of potential outcome, defined as an outcome had the subject followed a particular treatment, possibly different from the treatment he or she actually followed. In experimental or clinical settings, the individual-level causal effect of a treatment may be viewed as the difference in outcomes if a person had followed that treatment as compared to a placebo or a standard protocol (Morton and Williams 2010).

Consider, for example, a simple trial in which subjects receive either treatment a or a' . Suppose further that the subject is randomized to receive

treatments. This subject will have a single observed outcome Y that corresponds to the potential outcome Y under treatment a , denoted by $Y(a)$, and one unobservable potential outcome $Y(a')$, corresponding to the outcome under a' , because a subject cannot receive both treatments simultaneously, which is so-called *fundamental problem of causal inference* (Holland 1986). When we are interested in the individual causal effects of taking treatment a instead of treatment a' , the individual level causal parameter that could be considered is a subject's outcome under treatment a subtracted from his outcome under treatment a' , i.e., $Y(a) - Y(a')$. If, for a given subject, all potential outcomes are equal (i.e., Y does not depend on a), then for this subject, the treatment has no causal effect on the outcome. If the treatment has no causal effect on the outcome for any subject in the study population, we could say the causal null hypothesis holds. Thus, a fundamental problem with subject-specific causal effects is that they are difficult to identify, because it is difficult to observe the outcome under both a and a' without further data and assumptions, as in crossover designs without carryover effects (Piantadosi 2005, 515).

The concept of potential outcome allows us to see why treatment differences observed in RCT represent causal effect in this way. With a random binary indicator for treatment z_i (1 for treatment, and 0 for control), $y_{i1}(y_{i0})$ denote the potential outcome corresponding to $z_i = 1(0)$. The causal effect for *each* subject is $\Delta_i = y_{i1} - y_{i0}$, which is not observable, since only the potential outcome corresponding to the treatment actually received is observed. Thus, the causal treatment, or population-level, effect, $\Delta = E(\Delta_i)$, cannot be estimated by simply averaging the Δ_i 's. For an RCT, however, we can estimate by using the usual difference in the sample means between the two treatment conditions.

Let $n_1(n_0)$ denote the number of subjects assigned to the intervention (control)

group and let $n = n_0 + n_1$. If y_{ik} denotes the potential outcome of the i th subject for the k th treatment for the n subjects, we observe y_{ik} if the subject is assigned to the k th treatment condition ($k = 0, 1$). If $y_{i1}(y_{j0})$ represents the observed outcome for the i th (j th) subject in the $n_1(n_0)$ subjects in the intervention (control) group, we can express the observed potential outcomes for the n subjects as: $y_{i1} = y_{j1}$ with $i = j_1$ for $1 \leq i_1 \leq n_1$ ($y_{j0} = y_{j_0}$ with $i = j_0 + n_1$ for $1 \leq j_0 \leq n_0$). The sample means for the two groups and the difference between the sample means are given by

$$\widehat{\Delta} = \bar{y}_{\cdot 1} - \bar{y}_{\cdot 0}, k = \frac{1}{n_k} \sum_{i_k=1}^{n_k} y_{i_k k}, k = 0, 1. \quad (Eq. 1)$$

For an RCT, treatment assignment is independent of potential outcome, i.e., $y_{ik} \perp z_i$, where \perp denotes stochastic independence. By applying the law of iterated conditional expectation, it follows from the independent assignment that

$$E(y_{ik}) = E(y_{ik} | z_i = k) = E(y_{ik}), k = 0, 1. \quad (Eq. 2)$$

It then follows from (Eq. 1) and (Eq. 2) that

$$\begin{aligned} E(\widehat{\Delta}) &= \frac{1}{n_1} \sum_{i_1=1}^{n_1} E(y_{i_1 1}) - \frac{1}{n_0} \sum_{j_0=1}^{n_0} E(y_{j_0 0}) \\ &= E(y_{i_1 1}) - E(y_{j_0 0}) \\ &= E((y_{i_1} | z_i = 1) - E(y_{i_0} | z_i = 0)) \\ &= E(y_{i_1}) - E(y_{i_0}) \\ &= \Delta. \end{aligned} \quad (Eq. 3)$$

Thus, the difference between the sample means estimates the causal treatment effect in the RCT. The above shows that the approaches such as the two-sample t -test and regression models can be used to infer causal treatment effects in RCTs. Randomization is key to the transition from the individual level difference, $y_{1i} - y_{0i}$, to the sample means in (Eq. 1) in estimating the average treatment effect (ATE). However, for non-randomized trials, exposure to treatments may depend on the values of the outcome variable, in which case the difference between the sample means in (Eq. 1) generally does not constitute the average causal effect. Thus, associations found in observational studies generally do not constitute causal effect.

2. Observational Studies and Selection Bias

Selection bias is one of the most important confounders in observational studies. It is also called *pre-treatment confounders*, since selection bias is often caused by imbalance in baseline covariates before treatment assignment (King et al. 1994; Glynn and Quinn 2010). The potential-outcomes approach provides a framework for explicating the effect of selection bias. Consider an observational study with two treatment conditions and let z_i continue to denote the binary indicator of treatment assignment. If treatment assignment is not random, z_i may not be independent of the potential outcome. Thus, the condition $y_{ik} \perp z_i$ may not hold and the identity $E(\hat{\Delta}) = \Delta$ in (Eq.3) may fail, in which case $\hat{\Delta}$ no longer estimates the causal treatment effect Δ .

By considering treatment difference from the perspective of potential outcome, not only can we develop models to address selection bias, but also methods to provide degree of confidence for the causal relationship ascertained. An approach widely used to address selection bias is to include covariates as additional explanatory variables in regression analysis. However,

as in the case of explaining causation using randomization, such an approach does not have a theoretical justification, since without the potential outcome-based framework, it is not possible to analytically define selection bias. Another undesirable aspect of the approach is its model dependence, i.e., relying on specific regression models to control for the effect of confounding(Ho et al. 2007). For example, a covariate responsible for selection bias may turn out to be statistically insignificant simply because of the use of a wrong statistical model or poor model fit. Most important, despite such adjustments, some residual bias may remain due to our limited understanding of covariates for the relationship of interest and/or the limited covariates collected in most studies. Without being able to assess the effect of such hidden bias, it is difficult to interpret findings from such an *ad-hoc* approach.

3. Post-treatment Confounders in Randomized Controlled Trials

In RCTs, treatment assignment is independent of potential outcomes, and statistical models such as regression can be applied for causal inference. However, this does not mean that such studies are free from selection bias. In addition to pre-treatment selection bias, selection bias of different kind, treatment noncompliance and/or non-random dropout post randomization, is quite common in RCTs. For example, if the intervention in an RCT has so many side effects that a large proportion of subjects cannot tolerate it long enough to receive the benefit of the treatment, the ITT analysis is likely to show no treatment effect, even though those who continue with the intervention may benefit. Thus, we must address such downward bias in ITT estimates, if we want to estimate treatment effects for those who are either not affected by or able to tolerate the side effects.

4. Mediation Effect

Many social scientists are interested in how an intervention and/or moderation achieves its effect upon establishing the efficacy of the intervention. Mediation analysis helps answer such questions (Bullock et al. 2010; Bullock and Ha 2012). Suppose z_i is the indicator of treatment assignment, m_i the mediator, and y_i the outcome. By investigating such a mediational process through which the treatment affects outcomes, not only can we further our understanding of the underlying mechanisms, but we may also develop alternative and better intervention strategies for the issues.

Structural equation models (SEM) are generally used to model mediation effects (Baron and Kenny 1986; Bollen 1987; Judd and Kenny 1981; McKinnon and Dwyer 1993). The mediation model illustrates how the treatment achieves its effect on the outcome y_i by first changing the value of the mediator m_i . For a continuous m_i and y_i , the mediation effect can be modeled by the following SEM:

$$\begin{aligned} m_i &= \beta_0 + \beta_{zm} z_i + \epsilon_{mi}, \\ y_i &= \beta_1 + \beta_{zy} z_i + \beta_{my} m_i + \epsilon_{yi}, \quad \epsilon_{mi} \perp \epsilon_{yi} \quad (\text{Eq. 4}) \end{aligned}$$

In this framework, the parameter β_{zy} is interpreted as the direct effect of treatment on the outcome y_i , while $\beta_{zm} z_i$ is interpreted as the indirect, or mediated, effect of the treatment z_i on the outcome y_i through m_i . Thus, the total effect of treatment is viewed as the combination of the direct and indirect effects, $\beta_{zy} + \beta_{zm} \beta_{my}$.

The SEM overcomes the limitations of standard regression models to accommodate variables that serve both as a dependent and independent variable such as the mediator m_i . However, since it is still premised upon

the classic modeling paradigm, it falls short of fulfilling the goal of providing causal effects. Causal inference for mediation analysis can also be performed under the paradigm of potential outcomes. Note that the error terms m_i and y_i in (Eq. 4) are assumed independent. This sequential ignorability condition is critical not only for ensuring causal interpretation, but also for identifying the SEM in (Eq. 4) as well.

III. Statistical Tools for Causal Inference

In the presence of selection bias, not only models for cross-sectional data such as linear regression, but even models for longitudinal data such as mixed-effects models and structural equation models are not suitable for causal inference. A large body of literature has been accumulated to address this issue in both observational and RCT studies. One of the widely used approach is to view unobserved components of potential outcomes as *missing data* and employ the methodology for missing data to deal with the technical problems within the context of causal inference (Holland 1986; King et al. 1994; Rubin 2006). Thus, the goal of causal inference is now to model or impute the missing values, or the unobserved potential outcomes, to estimate the average causal effect $\Delta = E(y_{1i} - y_{0i})$, which is not directly estimable using standard statistical methods because of the fundamental problems of causal inference.

In practice, these issues are further complicated by missing data, and various approaches have been developed to address the issues. These models are largely classified into one of the two broad categories: parametric and non-parametric. Since the unobserved potential outcome can be treated as missing data, the parametric and non-parametric frameworks both seek to extend standard statistical models for causal inference by treating the latent

potential outcome as a missing data problem and applying missing data methods. If treatment assignment is not random, it may depend on the observed, or missing potential outcome, or both. If the assignment is completely determined by a set of covariates, \mathbf{x}_i , then the unobserved potential outcome is independent of treatment assignment once conditioned upon \mathbf{x}_i . This missing at random(MAR) assumption allows one to estimate the average causal effect $\Delta = E(y_{1i} - y_{0i})$. Thus, by identifying the unobserved potential outcome as a missing data problem, methods for missing data can be applied to develop inference procedures within the current context.

1. Estimating Causal Treatment Effects in Observational Studies

1) Case-Control Designs

Case-control studies are widely used to ascertain causal relationships in nonrandomized clinical studies. In a typical case-control study on the relationship between some exposure variable(such as smoking), and disease(such as lung cancer), researchers first select a sample from a population of diseased subjects. Such a population is usually retrospectively identified by reviewing the charts of patients' medical histories. Then, the next step is to select a sample of disease-free individuals, or controls, from a non-diseased population, with the same or similar socio-demographic and clinical variables, which are believed to predispose subjects to the disease of interest. Since the cases and controls are closely matched to each other in all predisposed conditions for the disease except for the exposure status, differences between the case and control groups should be attributable to the effect of exposure, or treatment. This approach can be justified from

potential outcomes framework. For example, if y_{i1} represents the outcomes from the case group, then the idea of case-control design is to find a control for each case so that the control's response y_{j0} would represent the case's unobserved potential outcome y_{i0} . Thus, we may use the difference $y_{i1}-y_{j0}$ as an estimate of the individual-level causal effect, i.e.,

$$y_{i1} - y_{j0} \approx y_{i1} - y_{i0}$$

Now, the computable sample average, $\bar{\Delta}_{cc} = \frac{1}{n_1} \sum_{i=1}^{n_1} y_{i1} - \frac{1}{n_0} \sum_{j=1}^{n_0} y_{j0}$, becomes a good approximation of the non-computable average $\bar{\Delta} = \frac{1}{n_1} \sum_{i=1}^{n_1} (y_{i1} - y_{i0})$, which is an estimate of the average causal effect Δ .

2) Matching and Propensity Score Matching

The case-control design reduces selection bias by matching subjects in the case and control group based on pre-disposed conditions. For the case-control design to work properly, however, we need to find good controls for the cases. If \mathbf{x}_i denotes the set of covariates for matching cases and controls, we must pair each case and control with identical or similar covariates. As the dimension of \mathbf{x}_i increases, however, matching subjects with respect to a large number of covariates can be quite onerous. A popular approach for matching subjects is the Propensity Score(PS) matching(Rubin and Rosenbaum 1986). This approach is premised upon the fact that treatment assignment dictated by \mathbf{x}_i is characterized by the probability of receiving treatment given the covariates \mathbf{x}_i , i.e.,

$$\pi_i = \pi(\mathbf{x}_i) = \Pr(z_i = 1 | \mathbf{x}_i). \quad (Eq. 5)$$

If \mathbf{x}_i is a vector of covariates such that $(y_{1i}, y_{0i}) \perp z_i | \mathbf{x}_i$, then we can show that

$$\Pr(\mathbf{x}_i | z_i = 1, \pi_i) = \Pr(\mathbf{x}_i | z_i = 0, \pi_i).$$

This shows that conditional on π_i , \mathbf{x}_i has the same distribution between the treated ($z_i=1$) and control ($z_i=0$) groups. Thus, we can use the one-dimensional Propensity Score in (Eq. 5), rather than the multi-dimensional and multi-type \mathbf{x}_i , to match subjects. For example, we may model π_i using logistic regression among others. With an estimated π_i , we can partition the sample by grouping together subjects with similar estimated propensity scores to create strata and compare group differences within each stratum using standard methods. Then, we can derive causal effects for the entire sample by weighting and averaging such differences over all strata. Although it is convenient to use and applicable to both parametric and semi-parametric models, the PS generally lacks desirable properties of formal statistical models such as consistency and asymptotic normality (King and Nielson Forthcoming).

Another major problem is that \mathbf{x}_i is only approximately balanced between the treatment groups, after matching or sub-classification using the estimated propensity score, especially when the observed covariates \mathbf{x}_i are not homogeneous in the treatment and control groups and/or one or more components of \mathbf{x}_i are continuous. Thus, this approach does not completely remove selection bias, although Rosenbaum and Rubin (Rosenbaum and Rubin 1985) showed otherwise via simulation.¹⁾ In addition, since the choice of cut-point for creating strata using the propensity score is subjective in sub-classification methods, different people may partition the sample

1) According to Rosenbaum and Rubin (1985), creating five sub-classes can remove at least 90% of the bias.

differently, yielding different estimates and even different conclusions, especially when the treatment difference straddles borderline significance. An alternative is to simply use the estimated propensity score as a covariate in regression analysis. This implementation is popular, since it reduces the number of covariates to a single variable, which is especially useful in studies with relatively small sample sizes. The approach is again *ad-hoc* and, like the parametric approach discussed above, its validity depends on assumed parametric forms of the covariate effects (typically linear).

3) Marginal Structural Models (MSM)

A popular alternative to PS is the marginal structural model (Hernán et al. 2002; Robins 1999; Hernán and Robins 2019 (forthcoming)), pioneered by Robins. Like PS, MSM uses the probability of treatment assignment for addressing selection bias. But, unlike PS, it uses the propensity score as a weight, rather than a stratification variable, similar to weighting in survey research (Horvitz and Thompson 1952; Blackwell 2012). By doing so, not only does the MSM completely remove selection bias, but also yields estimates with nice asymptotic properties. Another nice feature about the MSM is its readiness to address missing data, a common issue in longitudinal data (Hernán et al. 2002). Under MSM, we model the potential outcome as

$$E(Y_{ik}) = \mu_k = \beta_0 + \beta_1 k, 1 \leq i \leq n, k = 0, 1 \quad (Eq. 6)$$

Since only one of the potential outcomes (y_{i1} , y_{i0}) is observed, the above model cannot be fit directly using standard statistical methods. If treatment assignment is random, i.e., $y_{ik} \perp z_i$, then $E(y_{ik}) = E(y_{ik}|k)$ and thus

$$E(y_{ik,k}) = \beta_0 + \beta_1 k, 1 \leq i \leq n_k, k = 0,1 \text{ (Eq.7)}$$

Thus, for the RCT we can estimate the parameters $\beta=(\beta_0, \beta_1)^T$, including the average causal effect $\Delta=\beta_1$, for the model for the potential outcome in (Eq. 6) by substituting the observed outcomes from the two treatment groups in (Eq. 7).

For observational studies, z_i is generally not independent of y_{ik} . If \mathbf{x}_i is a vector of covariates such that $(y_{i1} + y_{i0}) \perp z_i | X_i$, then we can still estimate β by modeling the observed outcomes $y_{i,k}$ as in (Eq. 7), although we cannot use standard methods to estimate β and must construct new estimates. To this end, consider the following weighted estimating equations:

$$\sum_{i=1}^n \begin{pmatrix} \frac{z_i}{\pi_1} (y_{i1} - \mu_1) \\ \frac{1 - z_i}{1 - \pi_1} (y_{i0} - \mu_0) \end{pmatrix} = \mathbf{0}, \text{ (Eq.8)}$$

Although the above equation involves potential outcomes, the set of equations is well defined. If the i th subject is assigned to the first(second) treatment condition, then $i = i_1$ ($i = j_0 + n_1$) and $y_{i1}=y_{i11}(y_{i0}=y_{j_00})$ for $1 \leq i_1 \leq n_1(1 \leq j_0 \leq n_0)$. It follows that

$$\begin{pmatrix} \frac{z_i}{\pi_1} (y_{i1} - \mu_1) \\ \frac{1 - z_i}{1 - \pi_1} (y_{i0} - \mu_0) \end{pmatrix} = \begin{cases} \begin{pmatrix} \frac{1}{\pi_1} (y_{i_11} - \mu_1) \\ 0 \end{pmatrix} & \text{if } z_i = 1 \\ \begin{pmatrix} 0 \\ \frac{1}{1 - \pi_1} (y_{i_00} - \mu_0) \end{pmatrix} & \text{if } z_i = 0 \end{cases}$$

Thus, the estimating equation in (Eq. 8) are readily computed based on the observed data. Also, the set of estimating equation is unbiased, since

$$\begin{aligned}
 E\left(\frac{Z_i}{\pi_i} y_{ik}\right) &= E\left[E\left(\frac{Z_i}{\pi_i} y_{ik} \mid x_i\right)\right] \\
 &= E\left[\frac{1}{\pi_i} E\left(\frac{Z_i}{\pi_i} y_{ik} \mid x_i\right)\right] \\
 &= E\left[\frac{1}{\pi_i} E(Z_i \mid x_i) E(y_{ik} \mid x_i)\right] \\
 &= E[E(y_{ik} \mid x_i)] \\
 &= \mu_k.
 \end{aligned}$$

Thus, by the theory of estimating equations, estimates of β obtained by solving the estimating equations in (Eq. 8) are consistent.

2. Post-treatment confounders in Randomized Controlled Trials

The intention-to-treat(ITT) analysis compares the outcomes of subjects by randomized groups, ignoring treatment compliance and other deviations of study protocols. As a result, the ITT yields the effect of treatment confounded by all those violations. Despite being protected from pre-treatment selection bias through randomization, ITT estimates of treatment are typically downwardly biased, because of the “diluted” effect by post-treatment bias due to treatment noncompliance and/or missing data.

1) Estimation using Instrumental Variable

One way to address treatment noncompliance is to partition study objects into different types based on their impacts on causal treatment effects and then characterize the causal effects for each of the types of treatment

noncompliance (Angrist et al. 1993; Imbens and Rubin 1997; Imbens and Rubin 2015). One approach that has been extensively discussed in the extant literature is a partition of sample under study into four types of their compliance behavior.

1. Complier (CP): subjects compliant with assignment treatment (control or intervention).
2. Never-taker (NT): subjects who would take the control treatment regardless of what they are assigned.
3. Always-taker (AT): subjects who would take the intervention regardless of what they are assigned.
4. Defier (DF): subjects who would take the opposite treatment to their assignment

In practice, the DF generally represents a small proportion of the noncompliant group.

For the AT and NT group, $\Delta_i = y_{i1} - y_{i0} = 0$. Neither group contributes to causal effect. For DF, D_i is in the opposite direction of causal effect. Thus, only the CP subsample provides information for causal effect. Let $C_i = 1(0)$ if the i th subject is in the CP (otherwise). The causal effect for the CP group is

$$\Delta_{CP} = E(y_{i1} - y_{i0} | C_i = 1) \quad (Eq. 9)$$

The above is called the Complier Average Causal Effect (CACE). In contrast, the intention-to-treat (ITT) effect is given by $\Delta_{ITT} = E(y_{i1} - y_{i0})$.

If C_i is observed for each subject, then we have

$$\begin{aligned} \Delta_c &= E(Y_{i1}|C_{i1} = 1, z_1 = 1) - E(y_{i0}|C_{i1} = 1, z_1 = 0) \\ &= E(y_{i_{11}}|C_{i_{11}} = 1) - E(y_{i_{00}}|C_{i_{00}} = 1), (Eq. 10) \end{aligned}$$

where C_{ik} denotes the complier's status for the i th subject in the k th treatment group ($k=0, 1$). We can then estimate $E(Y_{ik} | C_{ik}=1)$ based on the Complier's subsample within the k th treatment condition using standard methods such as the sample mean. In practice, we can only observe compliance status D_{ik} for the assigned treatment condition. Although similar, D_{ik} is generally different from C_{ik} . For example, $D_{i1}=1$ includes both the CP and AT subsamples within the treated, while $D_{i0}=1$ includes the CP + NT subsample within the control condition. By conditioning on D_{ik} , we can estimate

$$\Delta_d = E(y_{i_{11}}|D_{i_1} = 1) - E(y_{i_{00}}|D_{i_0} = 1)$$

However, $\Delta_c \neq \Delta_D$ unless there are no AT nor NT subsample in the study population.

Let $p_1 = E(D_{i1}=1)$ and $p_0 = E(D_{i0}=1)$. Then p_1 represents the proportion of CP+AT in the intervention, while p_0 represents the proportion of AT+DF in the control condition. If we assume no DF, then p_0 becomes the proportion of AT and $p_1 - p_0$ represents the proportion of the CP group. Thus, we can express (Eq. 9) as

$$\Delta_{CP} = \frac{E(y_{i1} - y_{i0})}{p_1 - p_0} = \frac{\Delta_{ITT}}{p_1 - p_0}. (Eq. 11)$$

In other words, we can estimate the CACE by modifying the ITT estimate:

$$\hat{\Delta}_{CP} = \frac{\hat{\Delta}_{ITT}}{\hat{p}_1 - \hat{p}_0} = \frac{\bar{y}_{.1} - \bar{y}_{.0}}{\hat{p}_1 - \hat{p}_0}, \hat{p}_k = \frac{1}{n_k} \sum_{i_k=1}^{n_k} D_{i_k}, k = 0, 1.$$

The identity in (Eq. 11) depends critically on the assumption of randomization. This is because, to ensure that p_k has the aforementioned interpretation, we must have

$$p_k = E(d_{ik}) = \Pr(D_{ik} = 1, z_i = k),$$

which is only guaranteed under random treatment assignment. Because of the critical role played by z_i in identifying the CP in the presence of confounding by the AT and NT subsamples, z_i is called an instrumental variable(IV) and $\hat{\Delta}_{CP}$ is known as the IV estimate of CACE.

2) Estimation using Principal Stratification

The IV method is limited to binary compliance variables. A notable limitation of the IV is that its estimated treatment effect only pertains to a subgroup of compliers in the study population. In most real studies, compliance varies over a range of patterns. One popular approach for allowing for graded levels of treatment compliance is the Principal Stratification(PST)(Frangakis and Rubin 2002). The PST creates Principal Strata based on similar treatment compliance patterns and estimates causal effects within each Principal Stratum. In the special case of IV classification of noncompliance, PST provides estimates of treatment effect for each of the four groups, albeit only CP is of primary interest. By creating graded treatment compliance categories, PST provides a more granular relationship between exposure and treatment effects.

Let s_{ik} denote a categorical outcome that indicates levels of treatment compliance for the k th treatment condition and $\mathbf{s} = (s_{i1}, s_{i0})^T$. The basic principal

stratification P_0 is the set of distinct \mathbf{s}_i , i.e., $P_0 = \{p_i; 1 \leq i \leq L\}$, where L is the total number of principal strata and p_i is a collection of \mathbf{s}_j such that $\mathbf{s}_i = \mathbf{s}_j$ for $\mathbf{s}_i, \mathbf{s}_j \in p_i$, but $\mathbf{s}_i \neq \mathbf{s}_j$ for $\mathbf{s}_i \in p_i, \mathbf{s}_j \in p_m (i \neq m)$. A principal stratification P is a collection of sets that are unions of sets in the basic principal stratification P_0 . Thus, P is a coarser grouping of the distinct \mathbf{s}_i .

Consider, for example,

$$s_{ik} = \begin{cases} 1, & \text{if compliant} \\ 0, & \text{if noncompliant} \end{cases}$$

For each subject, the potential outcome of noncompliance status $\mathbf{s}_i = (s_{i1}, s_{i0})^T$ has four patterns, which constitutes the basic principal stratification: $P_0 = \{(1,1), (0,1), (1,0), (0,0)\}$. The four distinct patterns represent the CP (1,1), the DF (0,0), the AT (1,0), and NT (0,1) subsamples under the IV classification of treatment noncompliance. By combining some of the patterns in the basic principal stratification P_0 , we can create principal stratification P to represent noncompliance patterns of interest. For example, the principal stratification $P = \{(1,1), (1,0), (0,1), (0,0)\}$ no longer distinguishes between the CP and AT.

Once we establish an appropriate choice of principal stratification P , we can compare the potential outcome y_{i1} and y_{i0} within each P to define the causal effect of interest:

$$\Delta_l = E(y_{i1} - y_{i0} | \mathbf{s}_i), 1 \leq l \leq L,$$

The goal is to estimate Δ_l for each l th stratum. We may also create weighted averages to obtain overall treatment effects of interest. Inference about $\theta = \{\Delta_l; 1 \leq l \leq L\}$ can be based on maximum likelihood or Bayesian methods. In the special case of IV categorization, the PST provides more information about

the relationship between noncompliance and treatment effects than the IV. In addition to the CP, PST also provides treatment effects for the AT, NT, or even the DF group.

3) Estimation using Structural Mean Models

In most studies, there exists a large amount of variability in treatment noncompliance (Goetghebeur and Lapp 1997; Lui 2011). For example, again drawn from clinical studies, in a medication vs. placebo study, if the medication is prescribed daily for 2 weeks, exposure to medication can range from 0 to 14 days. We may group medication dosage using a graded categorical variable and apply the PST to characterize a dose-response relationship in this case. However, since this or any other grouping of the dosage variable is subjective, we may want to use the original number of days of medication use directly to more objectively characterize the dose-response relationship. Unfortunately, this will immediately increase the number of principal strata and may not provide reliable inference, or the PST may simply stop working, if there is not a sufficient number of subjects within every stratum. A more sensible approach is to treat such a continuous-like treatment compliance measure as a continuous variable to study treatment effect.

In many treatment research studies, active treatments are only available to study participants. In this case, active treatment is not available to the DF and AT subsample in the control condition, in which case causal treatment effect is determined by the AT+CP subsample in the treatment group. This allows us to model treatment effect as a function of continuous dose variables.

Let s_{ik} denote a continuous compliance variable for the i th subject in the k th treatment with $k = 1(0)$ for the active treatment(control) condition. For convenience and without the loss of generality, assume that $s_{ik} \geq 0$ with 0 representing zero dose. Since the active treatment is not available to subjects

in the control condition, $s_0 \equiv 0$ and thus s_0 provides no information about causal treatment effect. Thus we may model the causal effect as a function of s_1 only:

$$E(y_{i1} - y_{i0} | s_{i1}) = g(s_{i1}, \beta). \quad (Eq. 12)$$

where $g(s_1, \beta)$ is some continuous function of s_1 and β . Since $(y_1, y_0) \perp\!\!\!\perp Z$ for randomized studies, it follows that

$$\begin{aligned} \Delta(s_{i1}) &= E(y_{i1} | s_{i1}) - E(y_{i0} | s_{i1}) \\ &= E(y_{i1} | s_{i1}, z_i = 1) - E(y_{i0} | s_{i1}, z_i = 0) \\ &= E(y_{i1} | s_{i1}) - E(y_{i0} | s_{i1}, z_i = 0). \quad (Eq. 13) \end{aligned}$$

where i_1 again indexes the subjects assigned to the treatment group and y_{i1} is the observed outcome of the subject in the treatment group. The model in (Eq. 12) is the Structural Mean Model (SMM). To estimate $\Delta_i(s_{i1})$, we must evaluate $E(y_{i0} | s_{i1}, z_i = 0)$ so that it can be estimated with observed data. If s_1 is independent of y_0 , then we have

$$\begin{aligned} E(y_{i0} | s_{i0}, z_i = 0) &= E(y_{i0} | z_i = 0) \\ &= E(y_{i,0}) = \beta_0. \quad (Eq. 14) \end{aligned}$$

This *compliance non-selective* assumption is reasonable, if, for example, s_1 is not correlated with disease severity. In this case, (Eq. 13) is reduced to

$$\Delta_i(s_{i1}) = E(y_{i1} - y_{i0} | s_{i1}) = E(y_{i1} | s_{i1}1) - E(y_{i0}) = E(y_{i1} | s_{i1}1) - \beta_0. \quad (Eq. 15)$$

It then follows from (Eq. 12), (Eq. 13) and (Eq. 14) that

$$E(y_{i_0}) = \beta_0, E(y_{i_1}|s_{i_1}) = g(s_{i_1}, \beta) + \beta_0, 1 \leq i_k \leq n_k, k = 0, 1, n = n_0 + n_1. (Eq. 16)$$

Given a specific form of $g(s_{i_1}, \beta)$, the SMM in (Eq. 16) allows one to model and estimate treatment effects for continuous dose variables. For example, if $g(s_{i_1}, \beta) = s_{i_1} \beta$, the SMM has the form:

$$E(y_{i_0}) = \beta_0, E(y_{i_1}|s_{i_1}) = \beta_0 + s_{i_1} \beta_1, 1 \leq i_k \leq n_k, k = 0, 1,$$

or equivalently,

$$E(y_{i_1}|s_{i_1}) = \beta_0 + s_{i_1} z_{i_k} \beta_1, z_{i_k} = k, 1 \leq i_k \leq n_k, k = 0, 1,$$

Note that although s_{i_1} is missing for the control group, the above is still well defined, since $s_{i_1} z_{i_0} = 0$ for all $1 \leq i_0 \leq n_0$.

In many studies, we may collect sufficient information, say \mathbf{x}_i , to explain the compliance behavior s_{i_1} . In this case, we have

$$E(y_{i_0}|s_{i_1}, x_i, z_i = 0) = E(y_{i_0}|x_i, z_i = 0) = E(y_{i_0}|\mathbf{x}_{i_0}, z_{i_0} = 0). (Eq. 17)$$

Under this compliance explainable condition, the SMM can be expressed as

$$E(y_{i_1}|s_{i_1}, x_{i_1}, z_{i_1} = 0) = g(s_{i_1}, x_{i_1}, \beta) + (E(y_{i_0}|x_{i_0}, z_{i_0} = 0), z_{i_k} = k, 1 \leq i_k \leq n_k, k = 0, 1)$$

In the above medication vs. placebo studies, if treatment compliance is also tracked for the placebo group, then it is reasonable to assume that the variable

of placebo use, d_0 , explains treatment compliance, if the subject is assigned to the medication group. This is because under randomization subjects cannot distinguish between medication and placebo. Thus, if we let $x_0 = d_0$, then it follows that

$$E(y_{i_0} | x_{i_0}, z_{i_0} = 0) = E(y_{i_0} | d_{i_0}, z_{i_0} = 0)$$

We can readily model the right-side of the above. For example, if we model both x_{i_0} and d_{i_0} as a linear function, we have

$$E(y_{i_0} | d_{i_0}, z_{i_0} = 0) = \beta_0 + d_{i_0} \beta_1$$

We may also express the above in a compact form as

$$E(y_{i_k} | s_{i_1}, x_{i_k}, z_{i_k} = 1) = \beta_0 + \beta_{i_0} \beta_1 + z_{-i_k} s_{i_1} \beta_2, 1 \leq i_k \leq n_k, n_1 + n_0 = n.$$

As before, the above is still well defined, even if s_{i_1} is missing for the control group. In many intervention studies, the control condition offers either nothing or sessions that provide information unrelated to the intervention. In many studies, compliance can be tracked. However, a dose variable, d_0 , generally does not explain treatment compliance, if the subject is assigned to the intervention group, since the information disseminated through the control condition may have nothing to do with the information provided by the intervention condition. Subjects with high compliance in the intervention may be quite different from their counterparts in the control group. Thus, \mathbf{x}_i in the study should contain variables that help explain behaviors of compliance for the intervention.

3. Mechanisms of Treatment Effects

Understanding the causal pathways of treatment effect is critically important, since identification of causal mechanism not only improves our understanding of the issues of interest, but also allows one to develop alternative and potentially more effective and efficient intervention or prevention strategies. A popular approach for causal mechanism is mediation analysis.

1) Causal Mediation

In recent years, there has been heightened activities to develop models for causal mediation effect under the counterfactual outcome framework (e.g., Imai et al. 2010a; Imai et al. 2010b; Pearl 2001; Robins 2002; Robins and Greenland 1992).

Let m_{ik} denote the potential outcome of a mediator, m_i , for the i th subject corresponding to the k th treatment. The potential outcome of the primary variable of interest is more complex to allow one to tease out the direct and mediation causal effects of the intervention or exposure on this variable. Let $y_i(k, m_{ik'})$ denote the potential outcome of the variable of interest y_i corresponding to the k th treatment condition and mediator $m_{ik'}(k, k' = 0, 1)$. Note that in practice we can only observe m_{ik} and $y_i(k, m_{ik})(m_{ik'})$ and $y_i(k', m_{ik'})$, if the i th subject is assigned to the k th (k' th) treatment ($k, k' = 0, 1$). But, in order to extract the direct and mediation effects, we must consider $y_i(k, m_{ik'})$, which is not observed if $k \neq k'$ (Hafeman and Schwartz 2009; Pearl 2001). The direct effect of treatment is the effect of treatment, i.e.,

$$\zeta_{i(k)} = y_i(1, m_{ik}) - y_i(0, m_{ik}), \quad \text{for } k = 0, 1$$

This quantity $\zeta(k)$ is also called the natural direct effect or the pure(total) direct effect corresponding to $k=0(1)$. In addition, there is also controlled direct effect, $y(1, m) - y(0, m)$, which may be viewed as the treatment effect that would have been realized, had the mediator m_{ik} been controlled at level m uniformly in the postulation(Pearl 2001; Robins 2003; Robins and Greenland 1992). Note that $\zeta(1)$ is generally not same as $\zeta(0)$ and the difference represents interaction between treatment assignment and the mediator.

The causal mediation, or indirect effect, or natural indirect effect, is the difference between the two potential outcomes, $y(k, m_1)$ and $y(k, m_0)$, of the variable of interest resulting from the two potential outcomes of the mediator, m_1 and m_0 , corresponding to the two treatment conditions $k=1$ and $k=0$, i.e.,

$$\delta(k) = y_i(k, m_{i1}) - y_i(k, m_{i0}), \text{ for } k = 0, 1. \quad (\text{Eq. 18})$$

If the treatment has no effect on the mediator, that is $m_1 - m_0 = m$, then the causal mediation effect is zero. The quantity $\delta_i(0)(\delta_i(1))$ is also referred to as the pure indirect effect(total indirect effect)(Robins 1993). As in the case of direct effect, $\delta_i(1)$ and $\delta_i(0)$ are generally different.

The total effect of treatment is the sum of the direct and mediation effect:

$$\begin{aligned} \tau_i &= y_i(1, m_{i1}) - y_i(0, m_{i0}) = \delta_i + \zeta_i(0) \\ &= \frac{1}{2} [\delta_i(1) + \zeta_i + \delta_i(0) + \zeta_i(0)]. \end{aligned}$$

If we assume that no interaction between treatment assignment and the mediator, then $\delta_i(1) = \delta_i(0) = \delta_i$ and $\zeta_i(1) = \zeta_i = \zeta_i$. The total effect of treatment in this case is simply the sum of mediation and direct effect, i.e., $\tau_i = \delta_i + \zeta_i$.

In mediation analysis, we are interested in the Average Causal Mediation

Effect(ACME), $E(\delta(k))$, the average direct effect, $E(\zeta(k))$, and the average total effect, $E(\tau) = \frac{1}{2}[\sum_k(\delta(k)) + E(\zeta(k))]$. Under no mediator by treatment assignment interaction, the average total effect reduces to $E(\tau) = E(\delta) + E(\zeta)$.

2) Sequential Ignorability and Model Identification

The independence between the error terms in the SEM plays a crucial role in the causal interpretation of the mediation model. The pseudo-isolation condition plays a critical role for the identifiability of the parameters of the SEM. The issue of identifiability can be discussed under the potential outcome based inference paradigm. For example, Imai et al. has shown that if \mathbf{x}_i is a vector of pre-treatment covariates for the i th subject, then

$$z_i \perp \{y_i(k', m), m_{ik}\} | \mathbf{x}_i = \mathbf{x}, \quad k, k' = 0, 1, \quad (Eq. 19)$$

$$y_i(k', m) \perp m_{ik} | z_i = k, \mathbf{x}_i = \mathbf{x}, \quad k, k' = 0, 1.$$

The above is sequential ignorability(SI) because the first condition indicates that z_i is ignorable, given the pre-treatment covariates \mathbf{x}_i , while the second one states that the mediator m_{ik} is ignorable, given \mathbf{x}_i and the observed treatment assignment z_i . Although the first is satisfied by all randomized trials, the second is not. In fact, the second condition of SI cannot be directly tested from the observed data. Thus, sensitivity analysis is usually carried out to examine the robustness of findings under violations of the second ignorability assumption(Imai et al. 2010a).

Other assumptions have also been proposed. For example, Robins(Robins 2003) proposed the following condition for the identification of *controlled direct effect*.

$$\begin{aligned} z_i &\perp \{y_i(k', m), m_{ik}\} | \mathbf{x}_i = \mathbf{x}, & (Eq. 20) \\ y_i(k, m) &\perp m_{ik} | z_i = k, \mathbf{x}_i = \mathbf{x}, \mathbf{w}_i = \mathbf{w}, \end{aligned}$$

where \mathbf{w}_i is another set of observed post-treatment variables that confound the relationship between the mediator and the outcome. Under the more stringent assumptions, the following assumption is a necessary condition for identifying the ACME(Robins 2003):

$$y_i(1, m) - y_i(0, m) = B_i,$$

where B_i is a random variable independent of m . This is the so-called non-interaction assumption, which states that the controlled direct effect of treatment does not depend on the value of the mediator.

3) Models for Causal Mediation Effect

Under the stochastic independence in (Eq. 19), it can be shown that the ACME can be nonparametrically identified for $k = 0, 1$. Since the conditions in the SI imply $y_i(k', m) \perp z_i | m_{ik} = m', \mathbf{x}_i = \mathbf{x}_i$, it follows that for any k and k' :

$$E(y_i(k, m_{ik'}) | \mathbf{x}_i) = \int E(y_i | z_i = k, m, x_i) dF_{m_i | z_i=k', \mathbf{x}_i}(m). \quad (Eq. 21)$$

where $F_T(\cdot)$ ($F_{T|W}(\cdot)$) denotes the (conditional) cumulative distribution function (CDF) of a random variable T (T given W). We may further integrate out \mathbf{x}_i to obtain the unconditional mean:

$$E(y_i(k, m_{ik'})) = \int E(y_i(k, m'_{ik}) | \mathbf{x}_i) dF_{\mathbf{x}_i}(\mathbf{x})$$

By using (Eq. 21), we can derive direct, indirect and total effect for the linear SEM (LSEM) as the Generalized Linear Structural Equation Models (GLSEM), where m_i and y_i or both may be non-continuous variables (Vanstelandt and Goetghebeur 2003). For example, by expressing the LSEM in (Eq. 3) using the potential outcomes, we have

$$\begin{aligned} m_i(z_i) &= \alpha_1 + \beta_1 z_1 + \varepsilon_{i1}(z_i), \quad (\text{Eq. 22}) \\ y_i(z_i, m_i(z_i)) &= \alpha_2 + \beta_2 m_i(z_i) + \gamma z_i + \varepsilon_{i2}(z_i, m_i(z_i)), \end{aligned}$$

Note that to indicate the dependence of the potential outcome of the mediator m_i as a function of treatment assignment, we use $m(z_i)$, rather than m_{ik} , in the LSEM in (Eq. 22). The first condition in (Eq. 19) implies

$$E(\varepsilon_{i1}(z_i)|z_i = k) = E[\varepsilon_{i1}(z_i)] = 0,$$

while the second indicates that

$$E[\varepsilon_{i2}(z_i, m_i(z_i))|m_i = m, z_i = k] = E[\varepsilon_{i2}(k, m)] = 0.$$

It then follows that

$$\begin{aligned} E[y_i(k, m_{ik'})] &= E_{m_i|z_i=k'} [E_{y_1}|m_i = m, z_i = k] \\ &= E_{m_i|z_i=k'} [\alpha_2 + \beta_2 m_i + \gamma E(z_i = k)] \\ &= \alpha_2 + \beta_2 E[(\alpha_2 + \beta_1 z_1)|z_i = k'] + \gamma E(z_i = k). \end{aligned}$$

The ACME for the LSEM in Eq. 22 is

$$\begin{aligned}
 E(\delta(k)) &= E[y_{i(k, m_i(1))}] - E[y_{i(k, m_i(0))}] \\
 &= [\alpha_2 + \beta_2(\alpha_1 + \beta_1 E(z_i | z_i = 1)) + \gamma E(z_i = k)] \\
 &\quad - [\alpha_2 + \beta_2(\alpha_1 + \beta_1 E(z_i | z_i = 0)) + \gamma E(z_i = k)] \\
 &= \beta_2 \beta_1.
 \end{aligned}$$

Thus, under no mediator by treatment assignment interaction, the mediated effect is $E(\delta_1(1)) - E(\delta_1(0)) = \beta_2 \beta_1$, which is identical to the indirect effect derived from the classic LSEM in Eq. 4.

We can also obtain the different causal effect if there is no mediator by treatment assignment interaction. For example, if we assume an interaction of the form, $z_i m_i$, i.e.,

$$\begin{aligned}
 m_i z_i &= \alpha_1 + \beta_1 z_i + \epsilon_{i1}(z_i), \text{ (Eq. 23)} \\
 y_i(z_i, m_i(z_i)) &= \alpha_2 + \beta_2 m_i + \gamma z_i + \eta z_i m_i + \epsilon_{i2}(z_i, m_i(z_i)),
 \end{aligned}$$

Then by using arguments similar to the non-interaction case above, we obtain

$$\begin{aligned}
 E(z_i(k)) &= \gamma + \eta(\alpha_1 + \beta_1 k), \quad \text{(Eq. 24)} \\
 E(\delta_i(k)) &= \beta_1(\beta_2 + k\eta), \\
 E(\tau_1) &= \beta_2 \beta_1 + \gamma + \eta(\alpha_1 + \beta_1), k, k' = 0, 1.
 \end{aligned}$$

For the indirect (mediation), direct and total causal effect. These effects are again consistent with those derived from the classic LSEM approach.

The identification of ACME can be extended to the GLSEM. For example, if the mediator m_i is binary, but the outcome y_i is continuous, and m_i is modeled as: $E(m_i(z_i)) = \text{logit}^{-1}(\alpha_1 + \beta_1 z_i)$, where logit^{-1} denotes the inverse of the logit link function, then under no mediator by treatment assignment interaction it follows from (Eq. 21) that the ACME can be expressed as

$$\begin{aligned} E(\delta_1(k)) &= E[y_{i(k,m_{i1})} - y_i(k, m_{i0})] \\ &= \beta_2[\text{logit}^{-1}(\alpha_1 + \beta_1) - \text{logit}^{-1}(\alpha_1)]. \end{aligned}$$

Note that others have considered mediation analyses without using the SEM paradigm. For example, Rubin(Rubin 2004; Rubin 2005) and Jo et al.(Jo 2011) considered methods to estimate the causal effect of treatment in the face of an intermediate confounding variable(mediator) based on the framework of Principal Stratification. These methods are limited in their ability to accommodate continuous mediating and outcome variables and are less popular than their SEM-based counterparts.

4. Limitations and Problems of Potential Outcomes Framework

Although potential outcomes-based approach to causality has become very popular in statistics, social and health science, there are several practical limitations to this approach. One potential problem is that we may have to assume a continuous infinity of potential responses in complex problems where both treatment and outcomes are continuous and may vary in continuous time. Another issue is more fundamental and is related to the nature of the randomness of the potential outcomes: they vary in the population, but seem to be fixed for a subject. This has been well analyzed by Dawid(2000) who critiques this assumption as “a fatalist philosophy” and notes that “even after treatment has been taken, it seems unrealistic to regard the patient’s recovery status as predetermined.” One question is how to define causal effect without counterfactuals. The answer is to consider that different regimes can be applied to future patients, corresponding to different probabilities of treatment attribution. This is the approach taken by Arjas and Parner(2004), Geneletti and Dawid(2011) and Commenges and Gégout-Petit(2015). Rather

than split the response Y into several potential responses we can consider various regimes, giving rise to different probability laws for Y . This is not counterfactual because for instance an intervention will be made in the future for different individuals. The question is to infer from observational studies what will be the law of Y under this new set of treatments(regime).

IV. Discussion

Causal inference is widely used in social science to investigate the causal mechanism of exposures and interventions. Not only does research on this important issue have a long history, but the body of literature in this field is quite extensive as well, containing both methodological development and applications over a wide range of disciplines. The causal inference based on potential outcomes framework is by far most popular and plays a dominant role in the development of modern causal inference models and methods in social and biomedical sciences. This short essay aims to provide an overview of the relatively well-known methods for causal inference developed under the rubric of potential outcomes framework(or approach). The framework has many limitations, including many untested assumptions such as Stable Unit Treatment Value Assumption(SUTVA) and sequential ignorability. However, most of the popular statistical methods, such as the propensity score, principal stratification, marginal structural mean models and optimal dynamic treatment regime(Murphy 2003; Robins 2004) are developed based on this framework and has been proven quite useful. Using potential outcomes framework, these statistical methods can be generalized for causal inferences in various different settings.

References

- Arjas, E. and Parner, J.(2004). "Causal Reasoning from Longitudinal Data." *Scandinavian Journal of Statistics* 31: 171-187.
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects using Instrumental Variables(with Discussion)." *Journal of the American Statistical Association* 91: 444-472.
- Baron, Reuben M. and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51: 1173-1182.
- Blackwell, Matthew. 2012. "A Framework for Dynamic Causal Inference in Political Science." *American Journal of Political Science* 57(2): 504-520.
- Bollen, Kenneth. 1987. "Total, Direct and Indirect Effects in Structural Equation Models." *Sociological Methods* 17: 37-69.
- Brady, Henry E. 2008. "Causation and Explanation in Social Science." In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. New York: Oxford University Press.
- Bullock, John and Shang E. Ha. 2012. "Mediation is Harder than It Looks." In *Cambridge Handbook of Experimental Political Science*, eds. James N. Druckman, Donald P. Green, James H. Kuklinski and Arthur Lupia. New York: Cambridge University Press.
- Bullock, John G., Donald P. Green, and Shang E. Ha. 2010. "Yes, but What's the Mechanism?" *Journal of Personality and Social Psychology* 98(4): 550-558.

- Commenges, Daniel and Anne Gégout-Petit. 2005. "Likelihood Inference for Incompletely Observed Stochastic Processes: Ignorability Conditions." arXiv preprint math/0507151.
- Dawid, Alexander P. 2000. "Causal inference without counterfactuals." *Journal of the American Statistical Association* 95, 407-424.
- Frangakis, Constantine E. and Donald Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58: 21-29.
- Gary King and Richard Nielsen. Forthcoming. "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis*.
- Geneletti, Sara and Alexander P. Dawid. 2011. "Defining and identifying the effect of treatment on the treated." In Illari, P. M., Russo, F., and Williamson, J., eds. *Causality in the Sciences*. New York: Oxford University Press.
- Glynn, Adam N. and Kevin M. Quinn. 2010. "An Introduction to the Augmented Inverse Propensity Weighted Estimator." *Political Analysis* 18(1): 36-56.
- Goetghebeur, Els and Krista Lapp. 1997. "The Effect of Treatment Compliance in a Placebo-Controlled Trial: Regression with Unpaired Data." *Journal of Royal Statistical Society Series C* 46: 351-364.
- Hafeman, Danella M. and Sharon Schawartz. 2009. "Opening the black box: a motivation for the assessment of mediation." *International Journal of Epidemiology* 38: 838-845.
- Hernán, Miguel and James M. Robins. 2019(forthcoming). *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC.
- Hernán, Miguel, Babette A. Brumback, and James M. Robins. 2002. "Estimating the Causal Effect of Zidovudine on CD4 Count with a Marginal Structural Model for Repeated Measures." *Statistics in Medicine* 21: 1689-1709.

- Ho, Daniel, Kosuke Imai, Gary King and Elizabeth Stuart. 2007. "Matching as Non-parametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3): 199-236.
- Holland, Paul. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81: 945-970.
- Hollis, Sally and Fiona Campbell. 1999. "What is Meant by Intention-to-Treat Analysis? Survey of Published Randomized Controlled Trials." *BMJ* 319(7211): 670-674.
- Horvitz, Daniel G. and Donovan J. Thompson. 1952. "A Generalization of Sampling without Replacement from a Finite Universe." *Journal of the American Statistical Association* 47: 663-685.
- Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010a. "Replication Data for: A General Approach to Causal Mediation Analysis." *Psychological Methods* 15(4): 309-344.
- Imai, Kosuke, Luke Keele, and Teppei Yamamoto, T. 2010b. "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25: 51-71.
- Imbens, Guido and Donald Rubin. 2015. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.
- Imbens, GuidoW. and Donald B. Rubin. 1997. "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance." *The Annals of Statistics* 25: 305-327.
- Jo, Booil, Elizabeth A. Stuart, David P. MacKinnon, and Amiram D. Vinokur. 2011. "The Use of Propensity Scores in Mediation Analysis." *Multivariate Behavioral Research* 46(3): 425-452.
- Judd, Charles M. and David Kenny. 1981. "Process Analysis: Estimating Mediation in Treatment Evaluations." *Evaluation Review* 5: 602-619.

- King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- Lewis, David. 1973. *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lui, Kung-Jong. 2011. *Binary Data Analysis of Randomized Clinical Trials with Noncompliance*. Hoboken, NJ: Wiley.
- Lindberg, Steffan L., ed. 2009. *Democratization by Elections*. Baltimore, MD: The Johns Hopkins University Press.
- MacKinnon, David and James H. Dwyer. 1993. "Estimating Mediating Effects in Prevention Studies." *Evaluation Review* 17: 144-158.
- Mill, John Stuart. 1884. *A System of Logic: Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green, and Co., 1884.
- Morton, Rebecca B. and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. New York: Cambridge University Press.
- Murphy, Susan. 2003. "Optimal Dynamic Treatment Regimes." *Journal of Royal Statistical Society: Series B* 65(2): 331-366.
- Pearl, Judea. 2001. "Direct and Indirect Effects." In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, eds. Jack Breese and Daphne Koller. San Francisco, CA: Morgan Kaufmann: 411-420.
- Persson, Torsten and Guido Tabellini. 2005. *The Economic Effects of Constitutions*. Cambridge, MA: The MIT Press.
- Piantadosi, Steven. 2005. *Clinical Trials: A Methodological Perspective*. 2nd ed. Hoboken, NJ: Wiley.

- Robins, James M. 1994. "Correcting for Non-Compliance in Randomized Trials Using Structural Nested Mean Models." *Communications in Statistics: Theory and Method* 23: 2379-2412(1994)
- Robins, James M. 1999. "Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference." In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, eds. M. Elizabeth Halloran and Donald Berry. New York: Springer: 95-134.
- Robins, James M. 2003. "Semantics of Causal DAG Models and the Identification of Direct and Indirect Effects." In *Highly Structured Stochastic Systems*, eds. Peter J. Green, Nils Lid Hjort, and Sylvia Richardson. New York: Oxford University Press: 70-81.
- Robins, James M. and Sander Greenland. 1992. "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology* 3(2): 143-155.
- Robins, James. 2004. "Optimal Structural Nested Models for Optimal Sequential Decisions." In *Proceedings of the Second Seattle Symposium on Biostatistics*, eds. Danyu Lin and Patrik J. Heagerty New York: Springer.
- Rosenbaum, Paul R. 1984. "The Consequences of Adjustment for a Concomitant Variable that Has Been Affected by the Treatment." *Journal of Royal Statistical Society Series A* 147: 656-666.
- Rosenbaum, Paul R. and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41-55.
- Rosenbaum, Paul R. and Donald B. Rubin. 1985. "Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score." *The American Statistician* 39: 33-38.

- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66: 688-701.
- Rubin, Donald B. 1976. "Inference and Missing data(with Discussion)." *Biometrika* 63: 581-592.
- Rubin, Donald B. 2004. "Direct and Indirect Causal Effects via Potential Outcomes." *Scandinavian Journal of Statistics* 31: 161-170.
- Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of American Statistical Association* 100: 322-331.
- Rubin, Donald B. 2006. *Matched Sampling for Causal Effects*. New York: Cambridge University Press.
- Sekhon, Jasjeet S. 2008. The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods. In *The Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady and David Collier. Oxford University Press.
- Vansteelandt, Stijn and Els Goetghebeur. 2003. "Causal Inference with Generalized Structural Mean Models." *Journal of Royal Statistical Society Series B* 65: 817-835.

인과추론의 방법: 개념, 모델, 쟁점

이 병 재*

논문요약

인과추론은 사회과학의 가장 중요한 목표이다. 지난 30년간 통계학, 사회과학 및 의학 등의 분야에서 인과추론의 방법론에서 있어서 많은 개념적, 방법론적 진전이 이루어졌다. 이 논문은 인과추론 방법론의 기본적인 내용을 소개를 목적으로 하며, 반사실적 인과개념, 잠재적 산출 프레임워크 및 사회과학/의학에서 널리 사용되는 방법들에 대한 소개를 포함한다. 또한, 매개 분석 및 구조방정식 모형에 대한 인과모델 관점에서의 논의도 제시된다.

주제어: 인과추론, 성향점수 매칭, 주변구조모형, 인과매개분석,

투고일: 2019.02.15. 심사일: 2019.03.07. 게재확정일: 2019.03.12.

* 연세대학교 디지털사회과학센터, 연구교수.

[DOI] <http://dx.doi.org/10.21487/jrm.2019.03.4.1.151>

【연구경향】

Indeterminate Feature of Parameter Estimation in Multilevel Categorical Latent Variable

Jungkyu Park*

ABSTRACT

Structural indeterminacy among the multilevel discrete latent variables in the multilevel latent class model (MLCM) is discussed in this paper. Three scenarios - non-full-rank, independent, and permutation indeterminacy - are presented with theoretical explanations and proofs of each structural indeterminate case. Numerical examples are also included to provide intuitive and conceptual understanding of structural indeterminacy. The awareness of the structural indeterminacy in applying the MLCM to data is highlighted in the discussions. Researchers are giving examples and directions to check for problematic structures to ensure their final model has a theoretically sound latent structure when modeling data with multilevel discrete latent variables.

Keywords: Multilevel Latent Class Model, Model Indeterminacy, Model Identification

I. Introduction

Studies in psychology, education, and the social sciences have utilized latent variables to describe and conceptualize complex phenomena. Researchers

* Assistant Professor, Kyungpook University.

have captured latent variables and elegantly represented their constructs and components. A number of latent variable models have been proposed to establish a link between the observed variables and latent constructs, the best known of which is factor analysis(Jöreskog, 1969; Jöreskog and Sörbon, 1979; Gorsuch, 1983; Bollen, 1989). An important difference among these proposed models is the nature and assumed distribution of the latent components.

Factor analysis models postulate a continuous variable to describe the latent structure of the targeted concepts. However, latent class models(LCM) (Lazarsfeld and Henry, 1968; Goodman, 1974a, 1974b) hypothesized latent variables to be either discrete or categorical. Whether it is the nature of latent variables to be continuous or discrete has been debated, and several articles have focused on methods to differentiate these two opposite distributional assumptions(e.g., Haertel, 1990; Waller and Meehl, 1998; Muthén, 2006; Duncan, Stenbeck, and Brody, 1988; Steinley and McDonald, 2007). Although this differentiation has significance in terms of substantive interpretation, this paper limits the focus to cases in which discrete latent constructs are assumed.

The LCM utilizes a set of latent classes to explain observed relationships among the manifest variables. It is assumed that the relationships among manifest variables are explained by the subject's latent class membership and that different latent classes have specific patterns of probabilities in responding to items. Therefore, the LCM can be conceptualized as having several mutually exclusive and exhaustive internally homogeneous latent classes. Each subject belongs to one and only one latent class. In other words, all subjects in the same latent class have the same probabilities of responding to or endorsing a set of manifest variables. The elegance and simplicity of assumed categorical latent components offer an effective framework for many theories in the social sciences.

Vermunt(2003) proposed multilevel latent class models(MLCM) to take into

account the additional dependency induced by nested data(or multilevel data). He used random-effects approach(Bryk and Raudenbush, 1992; Goldstein, 1995; Snijders and Bosker, 1999; Rabe-Hesketh, Pickles, and Skrondal, 2001) to extend the LCM to adopt a multilevel nested data structure. He distinguished two types of random effects depending on the assumed distribution of the random effects. The first type of random effects follows a parametric continuous distribution. A normal distribution is usually hypothesized in order to facilitate parameter estimation. The second type of random effect is assumed discrete. In one example, multinomial distributed random effects replace normal distributed random effects in explaining the dependency in nested data. Discrete random effects have the advantage of weaker distributional assumptions and have less computational burden(Vermunt and Van Dijk, 2001). In practice, the discrete version of MLCM has been demonstrated to provide a good framework for application in different research fields. For example, Bijmolt, Paas, and Vermunt(2004) applied the framework to marketing studies and Dias, Vermunt, and Ramos(2010) applied it to a study in finance.

Discrete latent components offer an elegant representation for translating the research question into statistical terms. Moreover, hypothesized discrete latent quantities have been used to describe not only the nature of latent components at each level but also the dependency between levels. This paper discusses an important but often overlooked aspect of modeling data with multilevel discrete latent variables: indeterminacy. In LCM, indeterminacy usually relates to a model's identifiability, the issue of label switching, or decisions about the number of classes. However, in the context of MLCM, this paper concerns another aspect of indeterminacy, that is, structural indeterminacy. Specifically, this type of indeterminacy relates to relationships between latent components at different levels. Only a well-defined multilevel latent structure can offer valid and meaningful interpretation of the

relationships among the hypothesized discrete latent constructs. This paper illustrates three scenarios of structural indeterminacy in MLCM. These structural indeterminate scenarios will show that using multilevel discrete latent constructs to interpret relationships observed in data could be compromised.

The purpose of this paper is to draw attention to the potential harm caused by structural indeterminacy. Theoretical explanations and numerical illustrations are given to emphasize the problematic structure of MLCM. The remainder of this paper is structured as follows: First, the model specification and parameter estimation of MLCM are introduced, followed by discussions on selecting the number of latent components. Three scenarios of structural indeterminacy and their theoretical explanations of causing indeterminacy are presented, and numerical examples illustrate each scenario. The paper concludes with discussions and final remarks on the applications of MLCM in practice.

II. Formulation and Estimation of Multilevel Latent Class Models

In the MLCM discussed here, the latent variables at different levels are assumed discrete. Therefore, the models are similar to the nonparametric approach of the multilevel latent class model presented in Vermunt's(2003) paper. However, there are several conceptual differences between the two models. Instead of utilizing discrete random effects to account for the dependency, the approach here builds a LCM at each level. Specifically, one set of latent classes is used to explain dependency at a lower level(i.e., individual level), and another LCM is built to explain dependency at a higher

level(i.e., group level). In order to differentiate the components associated with different levels, categories of the discrete latent variable at the lower level are called latent classes, and categories of the discrete latent variable at the higher level are referred to as latent clusters. For formal specification of the MLCM, a classical LCM will be introduced and then extend it to incorporate multilevel latent structures.

Let X be the latent variable consisting of M mutually exclusive and exhaustive latent classes. The latent classes are indexed by m , where $m=1,2,\dots,M$. In addition, Y_{ij} denotes the random variable representing the response of subject i to the j th item($1 \leq j \leq J$), and y_{ij} represents a realization of the random variable Y_{ij} . The class-specific conditional probability of observing y_{ij} for item j of subject i in class m is $P(Y_{ij} = y_{ij} | X = m)$, which is also called the conditional response probability or latent conditional probability. In the LCM, the probability of obtaining the response pattern \mathbf{y}_i for subject i ($P(\mathbf{Y}_i = \mathbf{y}_i)$) is a weighted average of M class-specific conditional response probabilities, $P(\mathbf{Y}_i = \mathbf{y}_i | X = m)$:

$$\begin{aligned}
 P(\mathbf{Y}_i = \mathbf{y}_i) &= \sum_{m=1}^M P(X = m)P(\mathbf{Y}_i = \mathbf{y}_i | X = m) \\
 &= \sum_{m=1}^M P(X = m) \prod_{j=1}^J P(Y_{ij} = y_{ij} | X = m),
 \end{aligned}$$

where $P(X = m)$ denotes the probability of a randomly selected subject in the sample belonging to latent class m , which is usually called latent class probability.

Conceptually, latent class probabilities can be conceptualized as the class size of the latent class in the population. A LCM assumes local independence within each class so that the joint probability of response pattern \mathbf{y}_i in class

m is the product of all conditional probabilities of items within this class(line 2 of Equation(1)). A LCM also needs to satisfy the identification constraints. Goodman(1974a, 1974b) showed that latent class models are identifiable by the constraint $\sum_{m=1}^M P(X=m)=1$.

In order to identify and differentiate the clustering effects in different levels in MLCM, the nested data structure should be clearly represented in the notation system. Therefore, the observed response vector is now denoted as \mathbf{Y}_{ig} (instead of \mathbf{Y}_i) to indicate a subject's higher-level membership g (i.e., subject i is a member of group g).

A discrete random variable H_g with L possible outcomes is hypothesized to represent the L unique higher-level latent clusters. Each outcome of the discrete random variable can be conceptualized as a latent cluster consisting of homogeneous (higher level) groups sharing the same latent class distribution. This discrete random variable requires a parameter $P(H_g=l)$ to describe the distributions (i.e., cluster sizes) of the latent clusters. The differences between clusters are demonstrated in the latent class probabilities of each cluster. In other words, the latent class probabilities $P(X_{ig}=m)$ differ depending on the latent cluster's membership (H_g). They are represented specifically as $P(X_{ig}=m|H_g=l)$. The matrix $P(X|H)$ of size $M \times L$ is used to represent the estimated M latent class probabilities for the L clusters. The columns of the matrix $P(X|H)$ describe the within-cluster distributions of the latent classes. Accordingly, the entries of each column sum to one. Conceptually, each latent cluster can have a unique pattern of conditional response probabilities. Because different conditional response probabilities among clusters can be ambiguous and subject to questionable interpretations, the latent clusters are assumed to have no effect on the conditional response probabilities.

The parameters of MLCM can be estimated using the maximum

likelihood(ML) approach. The likelihood function is based on the probability density for the data of the higher-level unit. The log-likelihood to be maximized equals

$$\log L = \sum_{g=1}^G \log P(\mathbf{Y}_g | \mathbf{H}_g).$$

The ML estimates of the model parameters are obtained by a modified EM algorithm. The E-step modified by Vermunt(2003) is called the upward-downward procedure, which makes use of the conditional independence assumptions underlying the MLCM. The latent variables are summed out by going from lower- to higher-level units, and the marginal posteriors are subsequently obtained by going from higher- to lower-level units. The details of this algorithm are found in Vermunt(2003), and the outline of this algorithm is described in the Appendix.

III. Decisions on the Number of Latent Classes and Clusters

One important step of fitting empirical data to MLCM is to decide on the numbers of latent clusters and classes. This decision is the main research question in the task of model selection. There have been extensive studies on the issue of selecting the number of classes for LCM. For example, researchers have shown that popular chi-squared (χ^2) likelihood ratio statistics are not theoretically correct for model selections in LCM because the ratio between the two latent class models may not follow the χ^2 distribution if the reduced model($M-1$ classes) is obtained from the full model(M classes) by

placing parameters at their boundary values(Everitt and Hand, 1981; Aitkin, Anderson, and Hinde, 1981; Everitt, 1988). Various information-based criteria were suggested as alternative indices for selecting the number of classes for LCM.

The Akaike information criterion(AIC)(Akaike, 1973) and the Bayesian information criterion(BIC)(Schwarz, 1978) are probably the most common methods for model selection. The model with the lowest value of information criteria among several candidate models is considered the best-fit model. Several researches(e.g., Collins, Fidler, Wugalter, and Long, 1993; Lin and Dayton, 1997; Yang, 2006; Nylund, Asparouhov, and Muthén, 2007) have used simulation studies to systematically evaluate and examine the performance of several popular information criteria(e.g., AIC and BIC) and a few modified versions of information criteria(e.g., CAIC and adjusted BIC) on correctly uncovering the true numbers of latent classes in LCM.

Because there are two levels of latent components(i.e., latent classes and latent clusters) in MLCM, deciding the number of latent classes and clusters simultaneously becomes much more challenging. Moreover, the definition of “sample size” is unclear in the context of nested data structure. If the information criteria are used to decide the number of latent classes and clusters, the definition of “sample size” for some criteria that have the sample size as part of the penalty term(e.g., BIC and CAIC) is unclear in the context of nested data structure. For example, the sample size(N) in the formula can be the number of groups, the number of individuals, or the number of total sample size.

The task of simultaneous decisions on the number of both latent classes and clusters has received scant attention in the literature. However, Lukočienė and Vermunt(2010) conducted simulation studies to examine the performance of various model selection methods in deciding the number of higher-level

components. Lukočienė, Varriale, and Vermunt(2010) further extended the investigation on stimulating the performance of various criteria to determine the number of latent classes and clusters. They also proposed a stepwise model-fitting procedure to choose the number classes and clusters more effectively. These two papers shed light on deciding the quantities of latent components at either the lower level or the higher level. The indeterminacy relating to decisions on the number of latent classes and clusters is essential in MLCM. In this paper, however, the focus is on another type of indeterminacy: structural indeterminacy between the lower level and the higher level.

IV. Structural Indeterminacy among between-level Discrete Latent Components

Structural indeterminacy in MLCM concerns the relationships among discrete latent components between latent clusters and latent classes. Two definitions facilitate discussions of indeterminacy are given first, which are followed by three structural indeterminacy scenarios in MLCM. The theoretical explanations and proofs for each scenario are discussed separately.

Definition 1. An MLCM solution is said to achieve the optimum if the convergence criterion of the MLCM algorithm has been met.

For any given random start, $\{\hat{P}^{(0)}(H_g = l), \hat{P}^{(0)}(X_{ig} = m | H_g = l), \hat{f}^{(0)}(y_{ig}; \theta)\}$; the solution set $\{\hat{P}^{(t)}(H_g = l), \hat{P}^{(t)}(X_{ig} = m | H_g = l), \hat{f}^{(t)}(y_{ig}; \theta)\}$ obtained from the MLCM parameter estimation algorithm is said to arrive at the optimum if the algorithm terminates at the iteration t , which is within the maximally

allowable(pre-determined) number of iterations and is smaller than the convergence criterion.

Definition 2. An MLCM problem is identifiable(unambiguous or determined) if and only if its solution is optimal and unique from distinct random starts.

For distinct random starts, $\{\hat{P}^{(0)}(H_g = l), \hat{P}^{(0)}(X_{ig} = m | H_g = l), \hat{f}^{(0)}(y_{ig}; \theta)\}$; the MLCM is said to be identifiable if the estimation algorithm arrives at the same solution set $\{\hat{P}^{(t)}(H_g = l), \hat{P}^{(t)}(X_{ig} = m | H_g = l), \hat{f}^{(t)}(y_{ig}; \theta)\}$ within the maximally allowable number of iterations.

1. Non-full-rank Scenario

Let $P(X) \in R_+^{M \times 1}$ and $P(H) \in R_+^{L \times 1}$ denote the latent class probabilities and the latent cluster probabilities, respectively. The dependency of the latent classes and clusters in MLCM, that is, $P(X|H) \in R_+^{M \times L}$, are thus expressed as follows:

$$P(X|H) \stackrel{\text{def}}{=} \begin{bmatrix} P(X = m_1 | H = l_1) & \cdots & P(X = m_1 | H = l_L) \\ P(X = m_2 | H = l_1) & \cdots & P(X = m_2 | H = l_L) \\ \vdots & & \vdots \\ P(X = m_M | H = l_1) & \cdots & P(X = m_M | H = l_L) \end{bmatrix}.$$

The above matrix relates each latent cluster to its corresponding classes. According to the specifications of MLCM, the sum of the j th column of this matrix is 1 for all $j(j=1, \dots, L)$.

The number of solution sets of a linear system can be used to explain the indeterminacy of this scenario. The relationship between $P(X)$ and $P(H)$ can be related in terms of a linear system: $P(X|H)P(H) = P(X)$. From the theory of linear systems, the following relationships are known: 1) if the linear system

is consistent (i.e., the solution set is non-empty) and if the coefficient matrix $P(X|H)$ is full rank, then $P(X|H)P(H)=P(X)$ has a unique solution; 2) if $P(X|H)P(H)=P(X)$ is consistent but $P(X|H)$ is less than full rank, then $P(X|H)P(H)=P(X)$ has an infinite number of solutions. This implies that if $P(X|H)$ is not full rank, then the solution of the MLCM parameters is not identifiable since the solution set is not unique according to Definition 2. More specifically, when $P(X|H)$ is a non-full-rank matrix, then infinitely many $\hat{P}(H)$ that give $P(X|H)\hat{P}(H)=P(X)$ can be obtained.

2. Independent Scenario

The second case of structural indeterminacy occurs when the relationship between latent classes and clusters is independent. In other words, the compositions and sizes of each class are identical across all latent clusters. This independent scenario violates the fundamental assumption made in multilevel modeling, that is, the lower-level parameters depend on the status of higher-level parameters. Conditional independence cannot be tested using empirical data, but it has conceptual importance in statistical modeling. Without this dependency between levels, applying multilevel models to data may lead to invalid conclusions derived from explanations of dependency observed in the data.

Indeterminacy caused by the independent scenario can be formally explained by simple operations based on probability theory. When the relationships between a latent class m and a latent cluster l are independent of each other, the conditional probability describing the cross-level relationship is

$$P(X|H) = \frac{P(X \cap H)P(H)}{P(H)} = \frac{P(X)P(H)}{P(H)} = P(X).$$

Hence, the matrix describing the cross-level relationship $P(X|H)$ can be expressed as a matrix form by the concatenate L vector of $P(X)$ horizontally:

$$P(X|H) = [\underbrace{P(X) \cdots P(X)}_{L?}]_{M \times L}.$$

It is obvious that the cross-level relationship summarized as $P(X|H)$ in Equation(2) is now rank one. According to the non-full rank scenario discussed earlier, the independent scenario also results in a non-identifiable situation and thus leads to structural indeterminacy. Although this independent scenario can be included as a special case of *non-full-rank* scenario, it is discussed separately because the dependency between levels is the fundamental assumption in multilevel modeling. The independent scenario violates this assumption and thus jeopardizes the usages and applications of MLCM.

3. Permutation Indeterminacy Scenario

Redner and Walker(1984) first introduced the term label switching to describe invariance of likelihood under relabeling the mixture components. This problem is commonly noted in parameter estimation and clustering in the Bayesian approach, and is discussed in LCM and mixture models(e.g., Chung, Loken, and Schafer, 2004; Yang, 2006). Common suggestions to remedy this issue are to impose constraints on the parameter space(e.g., Stephens, 2000; Celeux, Hurn, and Robert, 2000). Within the context of MLCM, the problem of label switching can lead to structural indeterminacy, which is more

severe than classical LCM because it has two levels of latent components.

The simplest examples of this problem are when the posterior probability of observing an individual coming from each of the M classes is equal or when randomly selected groups from each L cluster are equiprobable. In such cases, the labeling of these classes and clusters is arbitrary. Another easy way to comprehend this indeterminacy is the concept of likelihood invariance: permutations of the latent components, that is, rearrangement of the component indices will not change the likelihood value.

The concept of permutation matrix can be utilized to describe this type of structural indeterminacy in MLCM. The function of a permutation matrix is to permute the rows and columns of a matrix, and a permutation matrix is formed by permuting rows or columns of an identity matrix to a desired order. One can find $M \times M$ permutation matrix \mathbf{R} ; pre-multiplying the matrix $P(X|H)$ with \mathbf{R} reorders the rows of $P(X|H)$. Similarly, another permutation matrix \mathbf{C} of size $L \times L$ can be defined; post-multiplying $P(X|H)$ by \mathbf{C} reorders the columns of $P(X|H)$.

The operation of pre- and post-multiplication $P(X|H)$ by permutation matrices \mathbf{R} and \mathbf{C} can be repeated many times. For example, if rows are permuted S times with $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_S$, and $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_T$ are used to permute columns $P(X|H)$ matrix T times, the configuration of the $P(X|H)$ after the permutation is $\mathbf{R}_S \dots \mathbf{R}_2 \mathbf{R}_1 P(X|H) \mathbf{C}_1 \mathbf{C}_2 \dots \mathbf{C}_T$. Some combinations of \mathbf{R}_s and \mathbf{C}_t exist that can permute the matrix back to the original configuration of $P(X|H)$. Specifically, this permutation process is

$$P(X|H) = \mathbf{R}_S \dots \mathbf{R}_2 \mathbf{R}_1 P(X|H) \mathbf{C}_1 \mathbf{C}_2 \dots \mathbf{C}_T.$$

Since the combinations of \mathbf{R}_s and \mathbf{C}_t are not unique, the elements of $P(X|H)$

can be permuted in different ways, but they can still permute back to a particular configuration. Note that a constraint is imposed on $P(X|H)$ in MLCM, that is, each column will sum to 1, so the choices of permutation matrix are not as free as in permuting rows and columns of a matrix without constraint. However, this constraint does not confine the combination of \mathbf{R}_i and \mathbf{C}_i as unique for securing structural determinacy in $P(X|H)$.

V. Numerical Examples and Illustrations

In this section, some numerical examples are used to illustrate the three indeterminate scenarios. These examples are designed to serve as intuitive examples to illustrate structural indeterminacy. Therefore, the easiest cases are used as examples. Note that these examples are for demonstration purposes only and many other general cases could be used for each scenario. In addition, because the MLCM is a probabilistic model, the discussions of indeterminacy here are situated at the conceptual level rather than in the context of numerical analysis.

The top panel of Table 1 illustrates examples of the *non-full-rank* scenario. Three numerical examples of different combinations in the number of latent classes(M) and clusters(L) are presented. The first two examples have identical distributions of the latent classes in both latent clusters, and the two-classes and three-clusters case have identical distribution in the first two columns but not in the third one. Simple values are chosen to illustrate this scenario; however, identifying the non-full-rank scenario in empirical data may not be easy or straightforward.

Numerical examples of the *independent scenario* are presented in the middle panel of Table 1. The first two examples are the equal-probable conditions

in which the probability of assigning a randomly sampled subject in each class is the same for both clusters. The next two examples illustrate fixed and identical ordering of class sizes for each latent cluster. Specifically, the likelihood of the assignment of a random sampled subject to a particular class is the same for both clusters.

Table 1. Numerical Examples of the Three Structural Indeterminate Scenarios

Examples of Non-full-Rank Scenario				
M/L	2/2	3/2	2/3	
$P(X H)$	$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/2 & 1/2 & 1/3 \\ 1/2 & 1/2 & 2/3 \end{bmatrix}$	
Examples of Independent Scenario				
M/L	2/2	3/2	2/3	3/2
$P(X H)$	$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$	$\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 1/3 \\ 1/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/2 & 1/2 & 1/3 \\ 1/2 & 1/2 & 2/3 \end{bmatrix}$	$\begin{bmatrix} 1/6 & 1/6 \\ 2/6 & 2/6 \\ 3/6 & 3/6 \end{bmatrix}$
Examples of Permutation Indeterminacy				
M/L	2/2	3/2	3/3	
$P(X H)$	$\begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$	$\begin{bmatrix} 1/4 & 2/4 \\ 1/4 & 1/4 \\ 2/4 & 1/4 \end{bmatrix}$	$\begin{bmatrix} 1/6 & 2/6 & 3/6 \\ 2/6 & 3/6 & 1/6 \\ 3/6 & 1/6 & 2/6 \end{bmatrix}$	

Examples for permutation indeterminacy are listed in the bottom portion of Table 1. These examples illustrate permutation indeterminacy when the patterns of the latent classes can be rotated into an identical pattern for each latent cluster. In the example of the two-classes and two-clusters model, one-third of individuals are categorized in latent Class 1 and two-thirds are in Class 2 for latent Cluster 1. However, the pattern of the two classes is reversed in Cluster 2: one-third of individuals are in latent Class 2 and two-thirds are in Class 1. Examples of permutation independency for the three-classes with two- and three-clusters case are also shown in Table 1.

VI. Discussions

The discussion of indeterminacy is not new regarding latent variable models. The factor indeterminacy was an important topic in the earlier development of factor analysis(Schönemann and Wang, 1972; McDonald, 1974, 1977; Steiger, 1979). A review of this topic is found in Steiger(1996). The LCM can be considered a discrete or categorical analogy to factor analysis. However, the discussion of indeterminacy within the traditional factor analysis framework is relatively more straightforward and obvious, but it is not as intuitive and apparent in the probability-based LCM. The earliest discussion on indeterminacy(unidentifiable) in latent class analysis appeared in Goodman's(1974) paper in *Biometrika*. However, since then few studies have focused on this issue when various models were extended from the LCM.

One possible reason for overlooking the issue of structural indeterminacy might be the lack of information about the estimation process. Usually, programs and software used for estimating MLCM do not automatically provide detailed information regarding the estimation process. There are no “warning” messages available in the output or log window when some problematic structures occur. Without information about problematic latent structure, users may easily miss checking for possible structural indeterminacy.

Moreover, different programs may have different routines to handle the problematic operation(e.g., non-full-rank). The solutions are usually based on the default settings of a specific software or program. For example, different kinds of software use different methods to handle the noninvertible matrix in the iteration process. The default settings of software are usually not familiar to users and are explained only in the technical manuals. Therefore, possible indeterminate structures may be masked by the default settings of the software.

When analyzing empirical data, researchers often rely on goodness-of-fit

tests or information criteria to identify the best-fit model among the candidate models. However, these tests and criteria are mainly good for selecting the “numbers” of latent components but not for detecting structural indeterminacy among components. Because there are no indices or tests to detect improper multilevel discrete latent structures, additional evaluations should be performed independently to check for possible structural indeterminacy in addition to making decisions on the number of latent components.

VII. Conclusion

To the best of my knowledge, since LCMs were extended to accommodate the nested data structure, no paper has discussed this important modeling concept in detail under a multilevel framework. This paper aimed to fill this gap by providing theoretical discussions and interpretations of structural indeterminacy in the context of multilevel discrete latent constructs. Three different scenarios: non-full rank, independent, and permutation indeterminacy were presented to illustrate structural indeterminacy in MLCM. Numerical examples corresponding to the three scenarios were also included to provide intuitive and conceptual understanding of the concept of structural indeterminacy.

In sum, a discrete version of MLCM has many advantages, among which is the potential use in empirical research. One example is its flexibility in translating hypothetical theories into the elements of a statistical model. However, the hypothesized multilevel latent structural model may be an indeterminate structure and thus lead to invalid interpretations of the final model. The awareness of this issue in applying MLCM to data is important in practice. The theoretical explanations of structural indeterminacy are

highlighted in this paper. The numerical examples provided give researchers starting templates and directions to check for problematic structures. I hope that the discussions of structural indeterminacy presented in this paper inspire and encourage researchers to check whether their final model has a theoretically sound structure when modeling data with multilevel discrete latent variables.

References

- Aitkin, M., Anderson, D., and Hinde, J. 1981. "Statistical modeling of data on teaching styles with discussion." *Journal of Royal Statistical Society, Series A* 144: 419-461.
- Aitkin, M., and Rubin, D. B. 1985. "Estimation and hypothesis testing in finite mixture models." *Journal of Royal Statistical Society, Series B* 47: 67-75.
- Akaike, H. 1973. "Maximum likelihood identification of Gaussian autoregressive moving average models." *Biometrika* 60: 255-265.
- Bijmolt, T. H., Paas, L. J., and Vermunt, J. K. 2004. "Country and consumer segmentation: Multi-level latent class analysis of financial product ownership." *International Journal of Research in Marketing* 21: 323-340.
- Bollen, K. A. 1989. *Structural equations with latent variables*. New York: Wiley.
- Bryk, A. S., and Raudenbush, S. W. 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Celeux, G., Hurn, M., and Robert, C. D. 2000. "Computational and inferential difficulties with mixture posterior distributions." *Journal of the American Statistical Association* 95: 957-979.
- Chung, H., Loken, E., and Schafer, J. L. 2004. "Difficulties in drawing inferences with finite-mixture models: A simple example with a simple solution." *The American Statistician* 58: 152-159.
- Collins, L. M., Fidler, P. F., Wugalter, S. E., and Long, L. D. 1993. "Goodness-of-fit testing for latent class models." *Multivariate Behavioral Research* 28: 375-389.
- Dias, J. G., Vermunt, J. K., and Ramos, S. 2010. "Mixture hidden Markov models in finance research." In Fink, A., Lausen, B., Seidel, W. and Ultsch,

- A. eds., *Advances in data analysis, data handling and business intelligence* pp. 451-459. Berlin-Heidelberg: Springer.
- Duncan, O. T., Stenbeck, M., and Brody, C. 1988. "Discovering heterogeneity: continuous versus discrete latent variables." *American Journal of Sociology* 93: 1305-1321.
- Everitt, B. S. 1988. "A finite mixture model for the clustering of mixed-mode data." *Statistics and Probability Letters* 6: 305-309.
- Everitt, B. S., and Hand, D. J. 1981. *Finite mixture distributions*. Chapman and Hall: London.
- Goldstein, H. 1995. *Multilevel statistical models*. 2nd edn. New York: Halstead Press.
- Goodman, L. A. 1974. "Exploratory latent structure analysis using both identifiable and unidentifiable models." *Biometrika* 61: 215-231.
- Gorsuch, R. L. 1983. *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Haertel, E. 1990. "Continuous and discrete latent structure models for item response data." *Psychometrika* 55: 477-494.
- Jöreskog, K. G. 1969. "A general approach to confirmatory maximum likelihood factor analysis." *Psychometrika* 34: 183-202.
- Jöreskog, K. G., and Sörbom, D. 1979. *Advances in factor analysis and structural equation models*. New York: University Press of America.
- Lin, T. H., and Dayton, C. M. 1997. "Model selection information criterion for non-nested latent class models." *Journal of educational and behavioral statistics* 22: 249-264.
- Lukočienė, O., Varriale, R., and Vermunt, J. K. 2010. "The simultaneous decision about the number of lower- and higher-level classes in multilevel latent class analysis." *Sociological Methodology* 40: 247-283.
- Lukočienė, O., and Vermunt, J. K. 2010. "Determining the number of components in mixture models for hierarchical data." In Fink, A. and Lausen, B. and Seidel, W. and Ultsch Eds., *Advances in data analysis*,

data handling and business intelligence pp. 241-249.
Berlin-Heidelberg: Springer.

- McDonald, R. P. 1974. "The measurement of factor indeterminacy." *Psychometrika* 39: 203-221.
- McDonald, R. P. 1977. "The indeterminacy of components and the definition of common factors." *British Journal of Mathematical and Statistical Psychology* 30: 165-176.
- Nylund, K. L., Asparouhov, T., and Muthén, B. 2007. "Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study." *Structural Equation Modeling* 14: 535-569.
- Rabe-Hesketh, S., Pickles, A., and Skrondal, A. 2001. "GLLAMM: A general class of multilevel models and a Stata program." *Multilevel Modelling Newsletter* 13: 17-23.
- Redner, R. A., and Walker, H. F. 1984. "Mixture densities, maximum likelihood and the EM algorithm." *SIAM Review* 26: 195-239.
- Schönemann, P. H., and Wang, M. M. 1972. "Some new results on factor indeterminacy." *Psychometrika* 37: 61-91.
- Schwarz, G. 1978. "Estimating the dimensions of a model." *Annals of Statistics* 6: 461-464.
- Snijders, T. A. B., and Bosker, R. J. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Steiger, J. H. 1979. "Factor indeterminacy in the 1930s and 1970s: Some interesting parallels." *Psychometrika* 44: 157-167.
- Steiger, J. H. 1996. "Coming full circle in the history of factor indeterminacy." *Multivariate Behavioral Research* 31: 617-630.
- Steinley, D., and McDonald, P. R. 2006. "Examining factor score distributions to determine the number of latent spaces." *Multivariate Behavioral Research* 42: 133-156.

- Stephens, M. 2000. "Dealing with label switching in mixture models." *Journal of the Royal Statistical Society, Series B* 62: 795-809.
- Vermunt, J. K. 2003. "Multilevel latent class models." *Sociological Methodology* 33: 213-239.
- Vermunt, J. K., and Van Dijk, L. 2001. "A nonparametric random-coefficients approach: The latent class regression model." *Multilevel Modelling Newsletter* 13: 6-13.
- Waller, N. G., and Meehl, P. E. 1998. *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.
- Yang, C. C. 2006. "Evaluating latent class analysis models in qualitative phenotype identification." *Computational Statistics and Data Analysis* 50: 1090-1104.

Appendix

MLCM Parameter Estimation Algorithm

Initialization: $t \leftarrow 0$, $\log L(\theta)_{\max} \leftarrow 10^7$

Input: $\hat{P}^{(0)}(H_g = l)$, $\hat{P}^{(0)}(X_{ig} = m | H_g = l)$, and $\hat{f}^{(0)}(y_{ig}; \theta)$

Compute $\log L^{(t)}(\theta)$

While $|\log L^{(t)}(\theta) - \log L(\theta)_{\max}| > 10^{-6}$ **or** $t \leq 10^5$ **do**

$\log L(\theta)_{\max} \leftarrow \log L^{(t)}(\theta)$ and $t \leftarrow t + 1$

E-step:

$$h_{igml}^{(t)} = \hat{P}^{(t-1)}(X_{ig} = m | H_g = l) \hat{f}^{(t-1)}(y_{ig}; \theta)$$

$$g_{ig'l}^{(t)} = \sum_{m=1}^M h_{igml}^{(t)}$$

$$\hat{X}_{igml}^{(t)} = h_{igml}^{(t)} / g_{ig'l}^{(t)}$$

$$\hat{H}_{gl}^{(t)} = \hat{P}^{(t-1)}(H_g = l) \prod_{i=1}^{n_g} g_{ig'l}^{(t)} / \sum_{l=1}^L \hat{P}^{(t-1)}(H_g = l) \prod_{i=1}^{n_g} g_{ig'l}^{(t)}$$

M-step:

$$\hat{P}^{(t)}(H_g = l) = \sum_{g=1}^G \hat{H}_{gl}^{(t)} / \sum_{g=1}^G \sum_{l=1}^L \hat{H}_{gl}^{(t)}$$

$$\hat{P}^{(t)}(X_{ig} = m | H_g = l) = \sum_{g=1}^G \sum_{i=1}^{n_g} \hat{H}_{gl}^{(t)} \hat{X}_{igml}^{(t)} / \sum_{g=1}^G \sum_{j=1}^{n_g} \sum_{m=1}^M \hat{H}_{gl}^{(t)} \hat{X}_{igml}^{(t)}$$

$$\hat{f}^{(t)}(y_{ig}; \theta) = \sum_{g=1}^G \sum_{i=1}^{n_g} \hat{H}_{gl}^{(t)} y_{ig} / \sum_{g=1}^G \sum_{i=1}^{n_g} \sum_{m=1}^M \hat{H}_{gl}^{(t)}$$

Compute $\log L^{(t)}(\theta)$

end while

다층 범주형 잠재변수에서 모수추정의 비확정성

박종규*

논문요약

본 논문에서는 다층 잠재계층모형에서 모수의 구조적인 비확정성에 대해서 다룬다. 먼저 다층 잠재계층에서 non-full-rank, independent, permutation indeterminacy의 세 가지 경우에 대해 소개하고 각각의 경우에 발생할 수 있는 모수의 비확정성에 대해서 설명한다. 또한 모수의 비확정성에 대한 이해를 높이기 위해 수리적인 예시를 제시한다. 논의에서는 다층 잠재계층모형을 이용해서 자료를 분석할 때 구조적 불확정성에 대한 인식이 중요하다는 점을 강조한다. 덧붙여, 연구자가 선택한 최종모형이 적절한 잠재 구조를 가졌는지를 점검하기 위한 지침을 제공한다.

주제어: 다층 잠재계층 모형, 모형 비확정성, 모형 식별

투고일: 2019.02.16. 심사일: 2019.03.08. 게재확정일: 2019.03.12.

* 경북대학교 심리학과, 조교수.