신뢰성 향상을 위한 설문문항 설계: 철도안전문화 설문자료 사례를 중심으로

임광균* · 김시곤**

Lim, Kwangkyun*, Kim, Sigon**

Designing Questionnaires for Better Reliable Survey: Case Study of Using Railway Safety Culture Survey Data

ABSTRACT

The study of safety culture estimate has received a wide attentions in many other industries, rather in the contexture of railroad stands at an initial phase. A factor analysis and internal reliability test were conducted to check if the survey results are reliable by looking at how the factors are dimensioned and how the questionnaires are sub-grouped in accordance with such factors, and further if there are any other common attributes being extracted as a new factor. The designed 5 factors were expanded to 7 through the test, and they were renamed accordingly. The questionnaires are reorganized according to the new 7 factors. It implicates that thorough investigation of the common attributes is important in reducing the gaps between interviewers' and interviewees' understanding. Moreover, the results of internal reliability test represent the new design of questionnaire is reasonably acceptable with the high reliable values. We expect such approaches can be applied for the railway culture survey study to be more reliable and applicable.

Key words : Railway safety culture, Internal reliability, Factor analysis, Factor extraction

초 록

철도외의 다른 산업분야에서 안전문화 측정에 대한 연구가 활발히 진행된 것과 달리, 국내 철도분야의 안전문화 성숙에 대한 연구는 아주 초보적인 단계에 머물러 있다. 본 연구는 철도안전문화 평가의 신뢰성을 높이기 위해 설문문항 구성 시 어떻게 평가요인을 구성하고, 설문문항 그룹이 해당 요인을 얼마나 잘 대변하고 있는지, 그리고 그룹을 정의한 기준이 다른 공통 속성을 갖는지 여부를 분석하기 위해 내적 신뢰도 및 요인 분석을 실시하였다. 요인분석 결과 기존 5개에서 7개 요소로 도출되었고, 그에 따른 설문문항의 그룹화 및 요소의 이름이 재부여 되었다. 이는 문항 간에 내재된 공통 속성을 잘 파악하는 것이 설문자와 피 설문자 간의 견해 차이를 줄이는 데 중요함을 의미한다. 또한 새로운 요인 분류에 대한 문항들의 내부 신뢰도 확인결과 여전히 이전과 비교하였을 때 높게 관측되어 그 분류가 적절함을 확인하였다. 본 연구는 철도안전문화 평가의 신뢰성을 높이는 데 통계적인 방법을 활용하여 기존 설문조사 결과를 검증하고 어떻게 접근해야 하는지 다루고 있어 관련분야에서 많이 활용될 것으로 기대된다.

검색어 : 철도안전문화, 내부신뢰도, 요인분석, 요인추출

* 정회원 · 교통안전공단 (Transportation Safety Authority · lim.kwangkyun@gmail.com)

** 정회원 · 교신저자 · 서울과학기술대학교 (Corresponding Author · Seoul Nat'l Univ. of Science and Technology · signonkim@seoultech.ac.kr)

Received June 25, 2014/ revised July 12, 2014/ accepted October 14, 2014

1. 연구의 배경 및 목적

안전문화는 한 기업의 안전수준을 나타내는 척도로서 안전성숙이라 불린다. Foster et al. (2013)은 한 기업의 안전문화 수준이 높아지기 위해서는 구성원의 자발적인 안전 이행의지와 책임감에 대한 문화인식 수준이 높아질 필요가 있으며, 아무리 훌륭한 안전향상을 위한 기술과 안전관리시스템이 뒷받침 된다 하더라도 적절한 수준의 안전문화 없이는 일정 목표 이상은 한계가 있다고 말한다. 지금까지 한국의 철도산업은 양적·기술적 측면에서 비약적인 발전을 이루었고 그러한 기술의 발전과 철도운영기관의 적극적인 안전수준 향상 노력 덕분에 철도사고 발생 또한 상당히 감소되었다. 아울러 한 층 더 발전된 철도안전 관리체계 구성을 위해 2012년 12월 철도안전법이 새롭게 개정되었다. 개정된 내용은 철도안전에 관련된 기술, 시스템, 인적·물적 자원, 직원 교육 및 자격 등에 대한 전반적인 관리를 효율적·신뢰적·투명하게 관리하도록 강조하고 있다.

철도안전관리체계의 구축은 분명히 철도의 종합적인 안전 향상에 크게 기여할 것임에 틀림없다. 그러나 어느 한계이상을 뛰어넘는 안정목표 달성을 위해서는 개별 구성원들의 안전 인식 및 태도가 그러한 시스템 향상과 더불어 반드시 성장되어야 한다. 회사 내 지속적인 안전에 대한 교육 및 사람들 간의 의사소통 증대, 상호간의 신뢰 구축 노력은 사람들의 안전 인식 및 태도 향상에 좋은 방법들이다. 그러나 그런 방법들을 적용하기 위해서는 가장 먼저 조직의 정확한 안전문화 수준이 진단되어야 하며, 그것을 바탕으로 안전문화 수준을 끌어올리기 위한 적절한 처방이 내려질 수 있다. 결론적으로 철도분야에서는 “어떻게 사람들의 인식 및 태도 수준을 평가할 수 있는가?” 라는 질문에 적합한 대답을 내릴 수 있는 관련연구가 미흡한 실정이다(Lim et al. 2014). 이는 안전문화 평가가 이미 정착되어 있는 해양산업(Fleming, 2000; Hudson, 2007), 원자력(IAEA, 2002), 항공(Gordon et al., 2007), 석유화학(Lardner, 2002; Filho et al, 2010), 광산(Foster and Houlst, 2013), 오일 및 가스 산업(Hudson, 2003) 분야와는 크게 대조를 이룬다.

이러한 관점에서 본 연구는 철도안전문화 성숙 평가의 신뢰성을 높이기 위해 설문구성 시 어떻게 평가 요인(여기서 요인이란 설문 문항을 그룹화 하는 그룹명을 말하며 주요인·보조요인 등으로 위계화 될 수 있음)을 구성하고, 각 요인별 문항들이 일관성을 나타내기 위한 방법이 무엇인지를 고민해 보았다. 실제로 적절한 절차(통계적 절차 포함) 없이 구성된 설문지를 통해 조사된 2012년 철도안전문화 설문 자료(TS 교통안전공단, 2012)를 활용하여 신뢰성(reliability) 및 요인분석(factor analysis)을 실시하였다. 신뢰성 및 요인분석은 요인 그룹과 각 요인별 설문문항 그룹이 얼마나 그 요인을 적절하게 설명하고 있는지 알아보기 위한 통계적인 방법이다. 분석을 통해

새로운 요인을 포함한 위계 제시 가능성을 검토하였다. 이러한 설문조사 사례를 통해 설문자의 최초 설문 기획 의도와 피설문자가 받아들이는 이해의 방향이 일부 다를 수 있음을 확인하는 것이 본 연구의 궁극적인 목적이다.

2. 기존 문헌 고찰

철도분야에서 안전문화 측정에 관한 연구는 아직까지 많이 활발하지는 않으나, 국내의 경우 한국철도공사와 같은 대규모 철도운영 회사를 중심으로 자체적인 안전문화 수준 측정을 최근에 실시하고 있는 것으로 파악되었다. 그러나 실제 어떤 설문구성으로 실행되었고 그 결과가 어떤지는 공개되지 않고 있다. 철도운영기관과는 달리 TS (2011, 2012)에서 2011년도 처음으로 ‘철도안전문화 측정기준 및 프로세스 개발 연구’가 진행되었다.

그 연구에 따르면 평가요소를 ‘철도안전문화지표’라 정의하고 5개 지표를 제시하였다: ‘안전정책 및 안전리더십, 이해당사자간 의사소통, 변화관리 및 자기진단, 학습문화, 안전책임마인드’. 첫 번째 요소(지표)인 “안전정책 및 안전리더십”은 9개의 설문문항, “이해당사자간 의사소통”에는 10개, “변화관리 및 자기진단” 8개, “학습문화” 9개, “안전책임마인드” 7개 문항 그룹들로 구성되었다.

실제 설문단계에서는 철도안전에 관련된 발생 가능한 6가지 시나리오를 제시하고 각 시나리오별 설문문항을 재배열하여 리커트 척도 1을 “매우 그렇다” 척도 5를 “매우 그렇지 않다”로 정의하여 측정하였다. 각 요소별 수준에 대한 결과는 해당 질문별 평가 점수에 대한 평균값을 사용하여 그 문화 수준을 나타내었다. 본 연구는 국내 철도안전문화를 처음으로 측정했다는 측면에서 커다란 의미가 있다. 그러나 설문구성을 위한 요소를 추출해 내고 어떻게 구성하여야 하는지에 대한 통계적인 검토는 전혀 고려되지 않았다. 즉, 비록 다른 산업분야에서 제시된 다양한 평가요소들이 반영되지 못하였더라도 연구에서 제시된 5개 요소로 철도안전문화가 적절히 설명되기에 충분하지, 그리고 요소별 해당되는 문항 셋이 적절하게 구성되었는지에 대한 통계적인 검증은 누락되었다. 이러한 적절한 평가요인의 발굴 및 이들의 위계 구성, 적합한 설문문항의 그룹화에 대한 사전평가 없는 설문조사는 그 결과의 신뢰성을 약화시킬 수 있다.

영국의 철도안전규제국(Office of Rail Regulation)은 평가요소 위계를 설정하여 전반적인 철도안전관리에 대한 성숙도를 RM3 (Railway Management Maturity Model) 라는 평가 틀을 개발하여 사용하고 있다(ORR, 2011). 그러나 안전성숙도를 동일하게 5단계로 표현하여 결과를 나타낸다 하더라도 설문에 의한 평가로 이루어지기 보다는 안전심사관들이 각 요소별 해당되는 증거자료들을 토대로 직접 평가를 내리는 절차로써 설문방식의 평가 방법과

는 근본적으로 다르다. 그러나 평가를 위한 평가요소들의 발굴 및 체계화는 오랜 현장 경험이 있는 전문가와 철도안전 심사관들이 함께 참여하여 만든 결과물이라는 측면에서 TS (2011, 2012)의 요인(지표) 구성 방법과는 대조된다.

이외에도 설문 신뢰성을 높이기 위한 측면의 연구를 살펴보면 Filho (2010)는 설문조사가 실행되기 전에 문항들의 적합성을 검증하기 위한 소규모 예비실험(Pilot test)을 강조하고 있다. 또한 실제 설문 조사 중에도 질문별 응답자의 일관된 응답이 결여 될 수 있으므로 이의 신뢰성을 높이기 위해 1주일 이내 동일 인물에 대해 다른 문항 구성(단 모든 문항의 내용은 동일)으로 재 설문 조사를 실시(동형 신뢰도 실험, Alternate-form reliability test)하도록 하고 있다(Lammers et al., 2005). 보다 구체적으로 Frazier et al. (2013)은 5개 서로 다른 산업분야의 안전문화측정에서 설문 요소의 위계 정립 및 각 요인별 설문문항 셋이 해당 요인을 얼마나 적절히 설명하고 있는지를 분석하여 설문요소의 위계를 재정립 하였다. 본 연구는교통안전공단에서 2012년 5개 평가 요소별 개발된 설문문항 셋이 실제로 그 평가 요소들을 얼마나 적절히 설명하는 지 신뢰성 및 요인분석을 통하여 설문문항 구성의 적절성을 확인하는 주목적으로 하였다.

3. 철도안전문화 설문조사 데이터

교통안전공단은 2011년부터 철도운영기관의 안전문화 수준을 매년 조사하여 왔다. 참고로 기존문헌 검토에서 언급된 바와 같이 본 연구는 어떠한 소규모 예비실험이나 응답의 신뢰성 확인을 위한 동형 신뢰도 실험은 실행되지 않았다. 이로 인한 설문 요인별 문항구성의 적절성 확인은 실제 설문조사 이전에 보완되지 않은 채 실시되었으나 본 연구에서는 이러한 부분은 무시하기로 한다. 또한 동형 신뢰도 실험 미 실시로 인한 피 설문자의 설문응답은 비록 1회 응답이라 하더라도 시간차를 둔 다수의 응답과 동일할 것이라고 가정한다(즉, ‘응답의 일관성 있음’으로 가정). 이러한 조건에서 실시된 설문조사 자료 중 2012년 10.29~11.7일 열흘 동안 수행된 자료를 활용하여 설문문항 간의 신뢰도 및 요인분석을 실시한다(2011년 자료는 미확보로 활용되지 못함). 설문은 10개 철도운영기관에 소속된 분야별(안전, 관제, 운행, 차량, 시설, 전기 및 신호) 안전담당자 및 직원 182명을 대상으로 실시되었고, 83명(응답률 45.6%)이 이에 응답하였다. 응답자의 60.2%는 40대 이며, 과장 및 차장 직급은 전체의 60.2%로 상당수가 철도분야 경력자로 구성되었다. 분야별로는 안전분야 종사자 42.2%, 시설분야가 16.9%를 차지하였다.

설문 조사된 분야(요소)를 살펴보면 TS (2011, 2012)는 HE (2005) 보고서에서 제시된 HMRI의 철도안전문화 구성요소를

기반으로 5개 요인 그룹을 정의하였다: ‘안전정책 및 안전리더십(경영진의 안전관리 가시성, 경영진이 승인한 안전정책의 안전문화 적정성, 경영진의 안전관련 마인드 및 직장 내 분위기 조성 등을 설명)’, ‘이해당사자간 의사소통(직원이 안전 논의 참여 여부, 직원의 안전문화에 대한 경영진에의 접근성, 직장내 신뢰 문화, 안전정보 전달 시스템, 안전정보의 이해도, 안전 우려사항의 기록, 조사, 완화 절차, 안전우려사항에 대한 내부 보고체계 등을 설명)’, ‘변화 관리 및 자기진단(직원의 변화관리 참여, 변화관리에 대한 직원 교육, 변화관리에 대한 직원 동기 부여 및 피드백에 대한 적극적 대응 등을 설명)’, ‘학습문화(사고조사 시스템, 사고에 따른 귀책 배정과정 및 징벌 절차, 사고원인에 대한 피드백 시스템 등을 설명)’, ‘안전책임마인드(안전문화/풍토 모니터링 시스템, 안전우선 행동, 성과 대 안전관리의 우선순위 평가, 이례사항 발생 시 안전책임의식 등을 설명)’. 각 그룹별 전체 43개의 설문문항으로 구성되어 있으며 실제 설문조사에 사용되었다.

4. 설문조사 분석

4.1 기술 통계(Descriptive statistics)

83명의 설문응답 자료를 토대로 각 질문별 기술통계 결과를 Table 1에 제시하였다. Q6은 43개의 질문유형 중 유일하게 반대로 응답(reverse-scored)되는 질문으로 작성되었다(Skenness<0, Q6. “경영진은 회사의 경영성과를 위해 안전보다는 회사의 이익을 우선 시하도록 직원을 압력 한다.”). 즉, ‘5. 매우 그렇지 않다’라고 답변한 경우가 긍정적인 안전문화를 의미한다. 전체적으로 상당수의 응답자가 자신이 속해 있는 조직의 안전문화가 긍정적이라고 평가하였다. 6개의 설문문항을 제외하고 대부분 분포의 중간 모양이 뾰족한 분포 형태를 갖는다(Kurtosis<0).

다음으로 43개의 문항은 철도안전문화를 표현하기 위한 하나의 그룹으로 그것들이 얼마나 그 그룹을 잘 표현하는지를 알아보았다. 이러한 것을 통계적으로 ‘내적 일치도(internal consistency)’라고 한다. 이 개념을 계산한 값이 Cronbach alpha(α)이며, 0과 1사이의 값으로 정의 된다(1에 가까울수록 내적 일치도가 높음). 철도안전문화를 설명하는 전체 43개의 질문을 모두 사용한 신뢰도 값은 0.92로 상당히 높게 측정되어 전체적으로 설문문항들이 철도안전문화를 효과적으로 설명하고 있는 것으로 보인다. 이러한 결과는 Frazier et al. (2013)의 $\alpha=0.95$ 값과 상당히 유사한 범위 내로 관찰되었다. 또한 5개 요인에 대한 내적 일치도 분석결과 모두 높게 측정되었다. 각각은: ‘안전정책 및 안전리더십’에 대해서 $\alpha=0.67$, ‘이해당사자간 의사소통’ $\alpha=0.78$, ‘변화관리 및 자기진단’ $\alpha=0.82$, ‘학습문화’ $\alpha=0.70$, ‘안전책임 마인드’ $\alpha=0.56$. 즉, 교통안전공단(2012)에서 시행된 철도안전문화 측정을 위한 5개 요인별 설문문항들은 각

Table 1. Descriptive Statistics

Q	N	Mean	Median	Std Dev	Skewness	Kurtosis
Q1	83	1.99	2	0.82	0.71	0.28
Q2	83	1.89	2	0.86	0.93	0.54
Q3	83	1.88	2	0.79	0.68	0.18
Q4	83	1.87	2	0.79	0.84	0.65
Q5	83	1.92	2	0.81	0.85	0.62
Q6	83	3.22	4	1.31	-0.18	-1.28
Q7	83	1.99	2	0.79	0.63	0.29
Q8	83	2.23	2	0.91	0.21	-0.80
Q9	83	1.98	2	0.83	1.11	1.98
Q10	83	1.93	2	0.85	0.87	0.43
Q11	83	2.13	2	0.96	0.92	0.76
Q12	83	1.64	2	0.73	0.87	0.17
Q13	83	1.48	1	0.65	1.56	3.52
Q14	83	1.92	2	0.83	0.82	0.41
Q15	83	1.82	2	0.70	0.70	0.85
Q16	83	1.58	2	0.54	0.15	-1.06
Q17	83	1.54	2	0.55	0.29	-1.03
Q18	83	1.80	2	0.69	0.52	0.07
Q19	83	1.95	2	0.81	0.65	0.14
Q20	83	1.73	2	0.73	1.03	1.48
Q21	83	1.95	2	0.82	0.76	0.34
Q22	83	1.77	2	0.74	0.77	0.49
Q23	83	1.87	2	0.89	1.21	2.08
Q24	83	2.12	2	0.97	1.24	1.97
Q25	83	2.07	2	0.92	1.20	2.10
Q26	83	2.01	2	0.83	1.40	3.24
Q27	83	1.77	2	0.67	0.55	0.39
Q28	83	1.90	2	0.88	1.08	1.37
Q29	83	1.90	2	0.84	1.21	2.18
Q30	83	1.81	2	0.86	0.97	0.42
Q31	83	1.75	2	0.85	1.12	0.80
Q32	83	1.99	2	0.94	1.19	1.54
Q33	83	1.92	2	0.90	1.09	1.24
Q34	83	1.77	2	0.85	1.33	2.30
Q35	83	1.77	2	0.75	1.28	3.39
Q36	83	2.37	2	1.00	0.78	0.29
Q37	83	1.76	2	0.77	1.26	2.93
Q38	83	1.53	1	0.57	0.49	-0.74
Q39	83	1.61	2	0.60	0.41	-0.65
Q40	83	1.87	2	0.79	0.69	0.15
Q41	83	1.64	2	0.62	0.73	1.25
Q42	83	1.70	2	0.62	0.62	1.10
Q43	83	2.19	2	0.92	0.67	0.24

해당 요인을 적절히 설명하는 문항들로 구성되었음을 의미하며, 결과적으로 요인 분석이 가능함을 뜻한다(즉, 문항들 간 공통개념을 설명할 가능성이 큼).

4.2 요인 분석(Factor Analysis)

앞서 언급된 높은 신뢰성(내적일치도) 값은 문항들이 해당 요인을 적절히 설명하는 것으로 판단되어 요인의 그룹과 그에 맞는 적절한 이름이 부여된 것으로 생각될 수 있다. 그러나 설문조사에서 동일한 개념을 측정하기 위해 설계된 리커드 척도 문항들이 정말 그런지 통계적 검증을 통해 알아볼 필요가 있다.

이를 위해서 설문문항들의 내재된 상관관계를 이용하여 요인을 구하고 이를 이용하여 문항들을 분류하고 그 그룹에 적절한 의미를 부여할 수 있는 요인분석이 도움이 된다. 물론 그 문항들의 내적 일치도는 Cronbach α 로 측정된다.

요인분석이란 n 개의 설문문항들이 상호 어떤 관계가 있는지 결정하여 문항들을 m 개의 문항 그룹으로 나누는 것을 의미하며, 이는 문항에 내재된 관계를 설명할 상호 독립된 요소를 얻고 몇 개의 요소로 내재된 관계를 충분히 설명할 수 있는가를 말한다. 이해를 돕기 위해 다음과 같은 예제를 생각해 보자. 물리와 수학이라는 두 가지 요인이 있고, 각 요인 안에는 이들을 적절히 설명하는 설문문항들이 구성되어 있다. ‘물리’와 ‘수학’이라는 이름은 분명히 다른 영역으로 구분되나, 물리를 좋아하는 정도와 수학을 좋아하는 정도에는 분명히 어느 정도 양의 상관관계가 있을 수 있다. 즉, 두 과목 간에 ‘논리적인 문제해결’과 ‘수리적 사고’라는 속성은 공통적으로 존재한다. 결론적으로, 요인분석은 이러한 무엇인가 공통적인 속성을 찾아내고 그에 따라 문항들을 재 그룹화 하는 것을 말한다.

설문문항의 그룹화는 요인 부하(loading)를 사용하고 그룹별 적절한 이름을 붙이는 것이다. 요인분석의 기본 가정에 따르면 각 설문문항은 m 개의 공통 잠재요소 f_m 에 가중치 1의 곱한 값과 설명되지 않는 오차 ϵ 의 합으로 나타낼 수 있다. 즉,

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} l_{11}l_{12}\dots l_{1m} \\ l_{21}l_{22}\dots l_{2m} \\ \vdots \\ l_{n1}l_{n2}\dots l_{nm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (1)$$

여기서, x_n 은 n 개의 설문문항이며, l_{nm} 은 n 번째 설문문항이 요소 f_m 에 미치는 가중치를 나타낸다. 이러한 가중치를 통계적으로는 ‘요인 적재량(factor loading)’이라 부르며 요인적재치가 어느 정도 크기 이상이어야 유의하다고 본다. 이에 대한 기준은 표본의 수, 설문문항의 수 및 요인의 개수 등에 따라 변동되며, 세부 설명은 아래 ‘요인 부하’에서 설명한다. 요인분석을 위하여 사회 및 행동과

학에서 광범위하게 사용되는 주 성분분석(Principal Component Analysis, PCA) 방법을 사용하였고, SAS 프로그램을 사용하였다. 비록 Ledesma and Valencia (2007)은 PCA 방법이 탐색적 요인분석(Exploratory Factor Analysis, EFA) 방법보다 요인 추출의 정확성 측면에서 신뢰성이 떨어진다고 언급하였으나 본 논문은 다른 요인분석 방법 간의 비교검토가 목적이 아니므로 실용분야에서 널리 활용되는 PCA 방법을 사용하였다.

4.2.1 공통요인 추출(Common Factor extraction)

PCA와 EFA 모두 다변량(Multivariate) 통계분석 방법이며, 중요한 것은 어떤 것을 적용하든 적절한 공통요소 f_m 을 추출하는 것이다. 예컨대, 과소추출(Under-extraction)은 잠재적인 모든 요인 그룹을 적절히 대변할 수 없고 이는 필요한 정보의 손실을 의미하며, 반면에 과대추출(Over-extraction)은 의미의 중복 및 해석의 어려움을 초래하여 모두 왜곡된 정보를 전달하게 된다(Tabachnik & Fidell, 2007). 지금까지 요소를 추출하는 다양한 방법들이 제안되었다. PCA와 마찬가지로 실무에서 가장 보편적으로 사용되는 방법은 Kaiser (1960)에 의해 개발된 K1 Rule (K1-Kaiser's eigenvalue-greater-than-one Rule)로서 eigenvalues (각 요인이 기존변수의 정보를 어느 정도 설명하는지를 나타내는 지표로서, 값이 크다는 것은 그 요인이 변수들의 분산을 잘 설명한다는 것을 의미함)가 1보다 큰 값을 공통요인으로 추출하게 된다. 몇 가지 문제점으로는 첫째, eigenvalue 값이 경계에 있는 경우 즉, 1.00과 0.99가 정말 서로 다른 결과를 갖기에 충분한가에 대한 논란이다. 둘째, 실제보다 과대 추출하는 경향이 있다. 이러한 문제점을 극복하고자 Cattell's Scree test (Cattell, 1966), Minimum Average Partial (Velicer, 1976) 테스트, Parallel Analysis (PA) (Horn, 1965)와 같은 대안들이 개발되었다. Zwick & Velicer (1986)는 특히 PA 방법이 가장 정확한 요인을 추출하는 것으로 보고하였다.

이러한 이슈에도 불구하고, 본 연구는 가장 보편적으로 사용되는 K1방법을 요인추출에 사용하였고 그 결과는 다음 Table 2와 같다.

Table 2. Factor Analysis Results: Eigenvalues and Percent of Variance

Factor	Eigenvalue	Variance percentage	Cumulative percentage
1	11.63	27.05	27.05
2	6.10	14.18	41.23
3	3.21	7.47	48.70
4	2.18	5.07	53.77
5	1.86	4.33	58.10
6	1.70	3.95	62.05
7	1.30	3.03	65.08

실제 분석결과 eigenvalue가 1보다 큰 요인이 9개로 도출되었으나, 다음 절의 요인 적재량 분석결과 8 & 9번째 요인에서 적재량 0.6 이상 해당되는 설문문항이 없어 그것들을 제외한 7개 요인을 최종 공통요인으로 추출하였다. 최종적으로 본 안전문화 설문조사의 27%가 Factor 1로 설명되며, 전체적으로 65%가 7개의 factor로 설명이 가능하다. 이는 Frazier et al.(2013)과 비교하였을 때 상대적으로 높게 관측되었다(50%).

4.2.2 요인 부하(Factor loading)

요인 부하 적재량의 절삭(Cutt-off) 기준은 많은 연구자들에 의해 제시되었다. 대표적으로 Tabachnik & Fidell (2007)은 0.33을 적재량의 최소 기준으로 제안하였고, Frazier et al. (2013)은 이러한 기준을 그대로 사용하였다. 그러나 이의 배경에는 설문문항 수 대비 응답자의 비율이 270배 이상으로 충분한 표본 수가 확보된 경우로서 상대적으로 훨씬 적은 표본 수(설문문항 대비 응답자의 비율 1.93배)를 사용한 본 연구의 경우 그 기준을 그대로 사용하기에는 무리가 있었다. 이에 Haier et al. (1998, p.112 참조)은 PCA 방법을 사용한 요인 부하 연구에서 표본 개수 별로 유의한 해석을 위한 적절한 절삭 기준을 다음 Table 3과 같이 제시하였고 본 연구에서는 0.6을 유의한 기준(응답자 83명)으로 사용하였다. 그러나 설문문항 대비 응답자의 비율이 1.93배로 아주 낮음에도 불구하고 이러한 기준이 본 연구의 경우에 그대로 적용될 수 있는지는 향후 추가적인 검증이 필요하다.

Table 4의 요인 부하(factor loading)결과, 절삭기준 0.6을 사용했을 때 43개중 31개 설문문항 만이 유의한 것으로 분석되었다(분석결과 최소 적재 값은 0.42임. 즉, 절삭 기준 0.4 사용 시 모든 설문문항 유의함). 예컨대, 설문문항 34, 5, 19, 33, 28, 40, 1, 13번은 모두 Factor1에 그룹화 될 수 있음을 의미한다. 12개의 누락된 설문문항의 경우 다른 설문문항과 그 의미가 유사한 경우가

Table 3. Significant Cut-off Points of Factor Loading by Sample Size

Cut-off Point	Sample Size needed for significance
0.30	≥ 350
0.35	250
0.40	200
0.45	150
0.50	120
0.55	100
0.60	85
0.65	70
0.70	60
0.75	50

Table 4. Factor Loadings

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Q34	0.870						
Q5	0.857						
Q19	0.788						
Q33	0.766						
Q28	0.720						
Q40	0.669						
Q1	0.661						
Q13	0.601						
Q20		0.859					
Q23		0.745					
Q30		0.722					
Q27		0.718					
Q4		0.716					
Q26		0.637					
Q24			0.867				
Q9			0.859				
Q25			0.823				
Q10			0.701				
Q11			0.699				
Q31				0.668			
Q35				0.663			
Q32				0.642			
Q43				0.631			
Q42				0.630			
Q36				0.614			
Q18					0.727		
Q17					0.706		
Q2						0.789	
Q29						0.605	
Q38							0.608
Q39							0.601

있었다. 예를 들면, 누락된 설문문항 ‘37. 회사 내 모든 직원의 안전에 대한 책무가 정의 되어 있다’ 와 ‘41. 자신의 역할과 책무를 이해하고 있다’는 모두 Factor 7의 ‘38. 본인과 다른 직원들의 안전에 대한 책임을 인식하고 있다.’ 와 비교하였을 때, 기본적인 역할과 책임 이해라는 측면에서 동일한 의미를 갖는 문항들로 분류된다. 마찬가지로 누락된 설문문항 ‘7. 경영진은 안전보다 회사의 이익을 우선시 할 경우, 야기되는 위험도를 이해하고 있다.’는 Factor 1의 ‘1. 경영진 및 관리자는 해당기관을 안전 시찰할 경우, 운영 구간의 위험도를 인식하고 있다.’와 위험도 인식이라는

측면에서 동일한 개념이다.

다음 Table 5는 요인 부하 결과를 토대로 문항들을 분류하고 그 그룹에 적절한 의미를 부여하고자 하였다. 이때, 그룹 내의 문항들이 공통적으로 의미하는 것을 따르되 다양한 산업분야에서 제시된 철도안전문화 성숙 평가 요소 이름을 최대한 활용하였다. 그러나 유사한 관점을 갖는 문항이 여러 그룹에 분류되거나 서로 다른 개념의 문항이 동일한 그룹에 묶여 있는 경우가 있어 명확히 그 그룹의 이름을 구분하기 어려웠다. 가령 Factor 1의 33번과 Factor 4에 속한 문항들은 사고조사라는 특징을 설명하는 공통개념

Table 5. Factor Description

Factor	Description	Quest. #
1	Safe system and interface	34, 5, 19, 33, 28, 40, 1, 13
2	Safety policy	20, 23, 30, 27, 4, 26
3	Management concern	24, 9, 25, 10, 11
4	Incident/Accident reporting and analysis	31, 35, 32, 43, 42, 36
5	Communication responsibility	18, 17
6	Management responsibility	2, 29
7	Personal Responsibility	38, 39

이며, Factor 5의 문항들은 전달확인과 그렇지 않을 경우의 책임에 대한 이해를 언급하는 내용으로 결국 Factor 7 ‘개인의 책임’과 일맥상통한다. 이러한 원인은 두 가지를 생각해 볼 수 있다. 첫째, 설문문항 대비 충분치 못한 표본(Table 2는 표본수를 사용한 기준이며, 설문문항 대비 비율은 미 고려됨). 둘째, 비록 앞선 요인 부하 분석을 통해 어느 정도 유사한 개념의 문항들이 이미 누락되었다 하더라도 연구자가 설문문항을 작성할 때부터 의도치 않게 서로 다른 그룹에서 동일한 개념을 갖는 다수의 문항(즉, 표현만 다름)이 여전히 존재 할 수 있다. 따라서 설문문항의 설계단계부터 영역 간 명확한 설문문항이 작성되기 위하여 문헌고찰에서 언급된 바와 같이 소규모 예비실험과 동형 신뢰도 실험은 필수적이다(Lim et al., 2014).

마지막으로 재분류된 그룹(요인)별 내부 신뢰성 분석 결과 Factor 1에 대해서 $\alpha=0.82$, Factor 2 $\alpha=0.71$, Factor 3 $\alpha=0.77$, Factor 4 $\alpha=0.51$, Factor 5 $\alpha=0.65$, Factor 6 $\alpha=0.71$, Factor 7은 $\alpha=0.75$ 으로 여전히 높은 내부 신뢰도를 보장하고 있다.

5. 결론 및 토론

5.1 결론

다양한 산업분야에서 각 분야의 특성에 맞는 안전문화 측정은 많이 활성화 되어 있다. 철도산업분야 또한 일부 철도운영회사 및 교통안전공단에서 철도안전문화 성숙도를 평가하기 위한 연구가 진행되어 왔으나, 대부분 단순히 평가지표(또는 요소)를 설정하고 그에 따른 설문문항을 작성하여 적용하고 있는 초보적인 단계에 머물러 있다. 즉, 평가의 신뢰성을 높이기 위해 설문구성 단계부터 고려해야할 평가요소 선택 및 검증, 예비실험, 해당 요소별 설문문항들의 요소에 대한 설명력을 확인하기 위한 통계적인 분석은 고려되지 못한 한계를 갖고 있다. 이러한 관점에서 본 연구는 2012년 철도안전문화 평가를 위해 실시된 데이터를 사용하여 요소별 설문문항들이 각각의 요인을 얼마나 잘 설명하는지에 대한

내적 일치도를 분석하였다. 이후 설문문항들의 내재된 상관관계를 이용하여 문항들이 기존에 분류된 요인에 맞게 적절히 분류되었는지 검증하기 위한 요인 분석을 실시하였다.

원래의 다섯 가지 요인의 설문문항 그룹에 대한 내적일치도 분석결과 ‘안전정책 및 안전리더십’에 대해서 $\alpha=0.67$, ‘이해당사자 간 의사소통’ $\alpha=0.78$, ‘변화관리 및 자기진단’ $\alpha=0.82$, ‘학습문화’ $\alpha=0.70$, ‘안전책임 마인드’ $\alpha=0.56$ 으로 대부분 높게 관측되었다. 즉, 각 설문문항 그룹이 해당 요인의 특성을 적절히 대변하고 있음을 의미한다. 그러나 설문조사에서 동일한 개념을 측정하기 위해 설계된 리커트 척도 문항들이 원래 주어진 속성 이외에 다른 공통적인 속성이 없는지 또는 내적 신뢰성을 더 높이기 위한 문항들의 재 그룹화가 필요한지에 대한 의문은 여전히 남는다. 이를 확인하기 위해 요인 분석을 실시하였고, 분석결과 두 개의 요인이 추가로 도출되었다. 설문문항은 요인부하 결과를 바탕으로 일곱 개 요인별로 재 그룹화 하고, 적절한 이름을 부여하였다. 새로운 일곱 개 요인에 대한 문항들의 내부 신뢰성은 이전과 비교하였을 때 여전히 높게 분석되었다(α 는 각 0.82, 0.71, 0.77, 0.51, 0.65, 0.71, & 0.75). 이것은 처음 설문문항 구성 체계와 실제 피 설문자가 받아들이는 이해의 방향에 차이가 있음을 보여준다. 향후 설문조사 시 문항 간에 내재된 공통 속성을 잘 파악하여 이에 맞는 그룹화와 적절한 그룹(요소)의 이름이 정해져야 이러한 차이가 줄어들 수 있을 것이다.

5.2 토론

본 연구는 설문문항 그룹이 해당 요인을 얼마나 잘 대변하고 있는지, 그리고 그룹을 정의한 기준이 다른 공통 속성을 갖고 있는지 내적 신뢰도와 요인 분석이라는 통계적인 방법을 활용하여 기존 설문조사 결과를 검증한다는 측면에서 의의가 있다. 그러나 연구결과와 정확성 및 활용성을 높이기 위해서는 다음과 같은 한계와 향후 추가적인 사항이 고려되어야 할 것이다.

- (1) 표본 수: 비록 본 연구는 Hair et al.(1998)이 제시한 절삭기준을 요인추출에 사용하였더라도 사용된 응답자의 수는 설문문항 대비 너무 적었다. 이것은 전체적인 통계분석 결과의 신뢰성을 저하시킬 우려가 있으며, 이것은 Table 5의 요인별 이름 재부여의 경우에서처럼 다른 그룹 간 동일한 개념을 갖거나 그룹 내 상반되는 문항들이 여전히 존재하여 그에 대한 어려움이 있었다. 즉, 표본의 수가 충분할 경우 하나의 공통 속성을 내포한 문항끼리 그룹 화될 가능성은 더욱 높아질 것이다.
- (2) 다른 그룹 간 문항들의 의미 유사성: 설문조사 문항은 처음부터 서로 다른 평가 분야로 분류되었음에도 불구하고 여전히 분야 간 동일한 개념을 갖는 문항들이 존재하였다. 비록 요인 부하의

결과가 이러한 동일한 개념을 갖는 문항을 식별하는데 조금이나마 역할을 한다 하더라도 궁극적으로 처음부터 잘못 설계된 문항들은 요인 부하의 해석을 어렵게 만들었다(즉, 그룹간의 경계 구분을 어렵게 함). 이러한 문제를 극복하기 위해 실제 설문조사 전 예비실험, 동형실험도 실험 등을 통하여 최대한 설문문항 구성의 정확성을 높여야 할 것이다.

- (3) 응답의 주관성: 본 연구는 리커드 척도로 응답된 결과를 사용하였다. 그러나 단순히 숫자로 표현된 응답은 비록 개인별로 생각하는 정도가 같다 하더라도 실제적으로 점수 부여에서 차이가 발생될 수 있다. 이의 대안으로 숫자로 표기된 리커드 척도 대신 상황을 구체적으로 기술한 상황 기술 답변(Unique descriptive answers) 방법이 응답자간 답변의 신뢰성을 높일 수 있을 것이다(ORR, 2011).
- (4) 다른 요인분석 방법의 비교: 본 연구는 일반적으로 가장 널리 쉽게 사용되는 PCA 방법과 K1기준을 적용하였다. 그러나 여러 연구에서 증명된 바와 같이 이러한 방법은 해석의 정확성과 신뢰성 측면에서 문제가 제기되었다. 따라서 분석방법 측면에서도 신뢰성을 확보하기 위해서는 EFA, PA방법과 같은 다른 개선된 통계분석 방법을 사용하여 추가적인 비교·검토가 필요할 것이다.

감사의 글

이 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다.

References

- Cattell, R. B. (1966). "The scree test for the number of factors." *Multivariate Behavioral Research*, Vol. 1, pp. 245-276.
- Filho, A., Andrade, J. and Marinho, M. (2010). "A safety culture maturity model for petrochemical companies in Brazil." *Safety Science*, Vol. 48, pp. 615-624.
- Fleming, M. (2000) Safety culture maturity model, Offshore Technology Report 2000/049.
- Foster, P. and Houlst, S. (2013). "The safety journey: Using a Safety Maturity Model for Safety Planning and Assurance in the UK Coal Mining Industry." *Minerals*, Vol. 3, pp. 59-72.
- Frazier, C., Ludwig, T., Whitaker, B. and Roberts, S. (2013). "A hierarchical factor analysis of a safety culture survey." *Journal of safety research*, Vol. 45, pp. 15-28.
- Gordon, R., Kirwan, B. and Perrin, E. (2007). "Measuring safety culture in a research & development centre: A Comparison of Two Methods in the Air Traffic Management Domain." *Safety Science*, Vol. 45, pp. 669-695.
- Hair, J. F., Tatham, R. L., Anderson, R. E. and Black, W. (1998). *Multivariate data analysis. (Fifth Ed.)* Prentice Hall, London.
- HE (Human Engineering). for the Health and Safety Executive, (2005) Development and validation of the HMRI safety culture inspection toolkit, Research Report 365.
- Horn, J. L. (1965). "A rationale and test for the number of factors in factor analysis." *Psychometrika*, Vol. 30, pp. 179-185.
- Hudson, P. (2003). "Applying the lessons of high risk industries to health care." *Qual Saf Health Care*, Vol. 12, pp. i7-i12.
- IAEA (2002). Safety culture in nuclear installations: Guidance for Use in the Enhancement of Safety Culture, IAEA-TECDOC-1329.
- Kaiser, H. F. (1960). "The application of electronic computers to factor analysis." *Educational and Psychological Measurement*, Vol. 20, pp. 141-151.
- Korea Transportation Safety Authority (TS) (2011). *Development of the safety culture inspection standards and process*, Technical Report (in Korean).
- Korea Transportation Safety Authority (TS) (2012). *Development of the safety culture inspection standards and process*, Technical Report (in Korean).
- Lammers, W. J. and Badia, P. (2005). *Fundamental of behavioral research*, California: Thomson and Wadsworth.
- Lardner, R. (2002). Towards a mature safety culture, Presented at the Institute of Chemical Engineers Annual Conference, Manchester, UK.
- Ledesma, R. D. and Valero-Mora, P. (2007). "Determining the number of factors to retain in EFA: An Easy-to-Use Computer Program for Carrying out Parallel Analysis." *Practical Assessment, Research & Evaluation*, Vol. 12, No. 2, pp. 1-11.
- Lim, K. K., Yeum, D. and Kim, S. (2014). "The development of a railway safety maturity model and estimate procedures." *Journal of the Korean Society of Civil Engineers*, Vol. 34, No. 1, pp. 195-202 (in Korean).
- ORR (Office of Rail Regulation). (2011). *Railway management maturity model (RM3)*.
- Tabachnik, B. and Fidell, L. (2007). *Using multivariate statistics (5th ed.)*, Needham Heights, MA: Pearson/Allyn & Bacon.
- Velicer, W. F. (1976). "Determining the number of components from the matrix of partial correlations." *Psychometrika*, Vol. 41, pp. 321-327.
- Zwick, W. R. and Velicer, W. F. (1986). "Comparison of five rules for determining the number of components to retain." *Psychological Bulletin*, Vol. 99, pp. 432-442.