

# Test-Retest Reliability of Paper-Pencil Test for Investigating Burner-Control Linkages of Four-Stove Gas Range

Dohyung Kee

Department of Industrial and Management Engineering, Keimyung University, Daegu 704-701

## ABSTRACT

**Objective:** The purpose of this study is to examine the test-retest reliability of paper-pencil test for investigating population stereotype of burner-control linkages of four-stove gas ranges. **Background:** Much of the research on gas-range stereotype strength has been performed using the paper-pencil test. While the use of the paper-pencil tests has the advantage of collecting data in a large group of subjects against reasonable costs, concerns regarding the reliability of the paper-pencil test has also been stated. **Method:** Three paper-pencil tests classified by intervals of test-retest were performed, in which eight types of burner-control arrangements were investigated. Sixty eight subjects were test-retested within two weeks, 85 subjects were test-retested within four weeks and 58 subjects were test-retested three times in series with interval of two and four weeks. **Results:** Chi-square test showed that three intervals of test-retest including two, four and two & four weeks have not statistically significant effect on subjects' responses irrespective of eight types of burner-control linkages dealt with in this study ( $p > 0.10$ ). Concordance rates of subjects' responses ranged between 73.5% and 100.0% for two-week test-retest, between 67.1% and 97.6% for four-week test-retest, and between 65.5% and 100.0% for two- and four-week serial test-retest. The concordance rates were linearly correlated with proportion of subjects' responses for stereotypes of burner-control linkages. **Conclusion:** It is concluded that the paper-pencil tests for investigating gas range relevant stereotypes could reproduce reliable results compared to baseline test within interval of four weeks. **Application:** The results of this study would be useful as an ergonomic guideline when designing the paper-pencil tests for stereotype relevant studies.

Keywords: Stereotype, Gas range, Burner-control linkage, Test-retest reliability

## 1. Introduction

4구 가스렌지 버너-조종장치 스테레오타이프는 인간 공학의 양립성 분야의 주요 연구 주제로, Chapanis와 Lindenbaum(1959) 이후 많은 연구가 수행되어져 왔다. 현재까지 수행된 스테레오타이프에 대한 많은 연구는 설문조사 방식으로 수행되어 왔다(Hoffmann, 2009). Shinar와

Acton(1978), Hsu와 Peng(1993), Wu(1997), Kee (2011a, b) 등은 설문조사를 이용하여 4구 가스렌지 스테레오타이프를 연구하였다.

설문조사는 실행이 용이하고 작은 비용으로 많은 참여자를 대상으로 다양한 요인에 대한 조사가 가능한 장점이 있으나, 설문조사 자체의 신뢰성 혹은 재현성(reproducibility)의 문제가 제기되어 왔다(IJmker et al., 2008). 설문조사법의 신뢰성은 컴퓨터 작업 자세(IJmker et al., 2008), 요통과

Corresponding Author: Dohyung Kee, Department of Industrial and Management Engineering, Keimyung University, Daegu 704-701.

Mobile: +82-10-5656-1260, E-mail: dhkee@kmu.ac.kr

Copyright©2013 by Ergonomics Society of Korea(pISSN:1229-1684 eISSN:2093-8462). All right reserved.

©This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. <http://www.esk.or.kr>

같은 근골격계질환 유발 요인(Coluci et al., 2009; Halpern et al., 2001; d'Errico et al., 2007), 작업 부하 변화(Miyake, 2009) 조사 등의 분야에서 다루어져 왔다. 위 연구에서는 설문조사를 일정 기간 간격으로 2번 혹은 3번 수행하였으며, 설문조사를 수행한 시간 간격은 1일, 1주, 2주, 7~10일 혹은 81일 내, 1년 및 5~6년 등으로 연구에 따라 다양하였다. 그러나 스테레오타이프 특히, 4구 가스렌지 스테레오타이프에 대한 설문조사법의 신뢰성을 다룬 연구는 현재까지 이루어진 적이 없다.

따라서 본 연구에서는 4구 가스렌지 버너-조종장치 연결 형태에 대한 스테레오타이프를 조사하는 설문조사법의 신뢰성을 실험을 통하여 다루고자 한다. 본 연구의 결과는 설문조사를 이용하여 스테레오타이프를 조사하는 연구의 기초 설계 자료로 활용될 수 있을 것으로 기대된다.

## 2. Method

### 2.1 Subjects

본 연구에서는 설문조사법을 이용한 가스레인지 스테레오타이프 조사의 신뢰성 연구를 위하여 3회의 설문조사 연구를 실시하였다. 설문조사 연구는 조사 간격으로 구분되며, 기준 시점 및 2주, 기준 시점 및 4주, 그리고 기준 시점, 2주와 4주 간격 연속 실시로 구분되며 총 7회의 설문조사가 수행되었다. 2주 간격 설문조사에는 남녀 대학생 68명(남 60명, 여 8명, 연령 평균 21.2세, 표준편차 1.61세, 범위 19~25세), 4주 간격 조사에는 남녀 대학생 85명(남 77명, 여 8명, 연령 평균 21.8세, 표준편차 1.75세, 범위 19~27세)이 참여하였다. 1주, 2주 및 4주 간 연속 3회 실시한 설문조사에는 남녀 대학생 58명(남 50명, 여 4명, 연령 평균 21.4세, 표준편차 1.66세, 범위 19~25세)이 참여하였다.

### 2.2 Experimental apparatus

본 연구에 사용된 설문지는 성별, 연령 및 본인 식별을 위한 문항 등 피실험자 인적 사항을 묻는 3문항과 4구 가스렌지 버너-조종장치 스테레오타이프에 대한 8문항 등 11개 항목으로 구성되었다. 스테레오타이프 조사 문항에서는 조종장치 아래에 있는 괄호 안에 각 조종장치와 연결되어 있을 것으로 생각되는 버너 기호를 기입하도록 하였다. 버너 표시에 숫자 혹은 알파벳 기호를 사용할 때 발생할 수 있는 암시효과(suggestive effect)를 최소화하기 위하여, Hsu와 Peng (1993)의 연구 결과를 따라 순서와 의미가 없는 기호 '#△□'를 사용하였다. 4구 가스렌지 스테레오타이프 조사 문

항은 일반적 가스렌지 형태와, 조사를 통하여 시중에 시판되고 있는 버너-조종장치 배열 형태 및 버너-조종장치 연결을 암시하는 암시 신호 등에 대한 8개로 구성되었다(Figure 1). 설문에 제시되는 가스렌지 모형은 단순한 선 및 원으로만 그려져 있는 기존 연구(Hsu and Peng, 1993; Shinar and Acton, 1978; Wu, 1997)와 달리, 3-D Max를 이용하여 최대한 실물에 가깝게 2.5차원으로 그려 현실감을 제고하였다.

### 2.3 Experimental design and procedure

실험 전에 실험 목적을 설명하고 실험 참여에 대한 동의를 구하였다. 실험에서는 설문지를 배부하고 인적 사항 3문항을 먼저 답하게 하고, 스테레오타이프에 대한 문항은 하나씩 모형 및 답변 방식을 설명한 후 답하게 하였다. 암시 신호가 있는 문항(Figure 1. (b)~(h))은 조종장치 위쪽 혹은 측면에 있는 암시 신호를 참조하여 답하고, 본인 가정에서 사용하고 있는 버너-조종장치 연결 형태가 아닌 직관적으로 연결되어 있을 것으로 생각되는 배열 형태로 답을 하도록 주의를 주었다. 앞서 언급한 바와 같이 두 번째 혹은 세 번째 설문조사는 첫 번째 설문조사 후 2주 혹은 4주 후에 실시하였으며, 이는 1~4주 간격으로 재현 설문조사를 실시하는 것은 수용할만 하다고 알려져 있기 때문이었다(Coluci et al., 2009). 또한, 두 번째 설문조사에서는 첫 번째 설문조사에



(a) range1



(b) range2



(c) range3

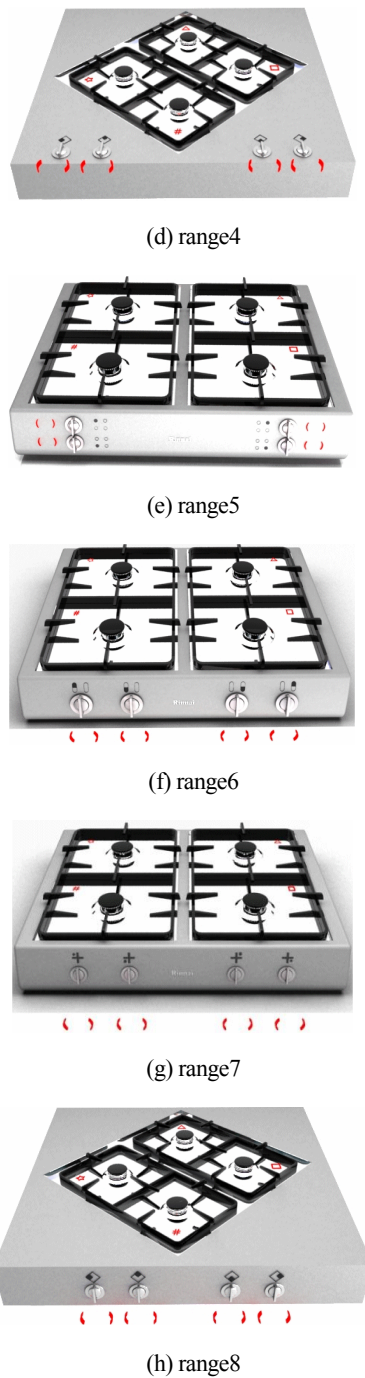


Figure 1. Gas range images used in the study

서 답한 내용을 떠올리려 하지 말고 현재의 직관에 의한 답을 하도록 하였다. 모든 설문조사는 피실험자 본인의 페이스로 답할 수 있도록 충분한 시간을 주었다.

### 3. Results

#### 3.1 Independence test

설문조사를 실시한 시간 간격이 피실험자의 응답에 영향을 미치는지를 알아보기 위하여 카이스퀘어(chi-square) 검정을 이용하여 독립성 검정(independence test)을 실시하였다(Table 1). Table 1에서 보는 바와 같이 8가지 가스렌지 형태에 관계없이 설문조사를 실시한 간격 2주, 4주 및 2, 4주 연속 실시가 버너-조종장치 연결에 대한 피실험자 응답 비율에 통계적으로 유의한 영향을 미치지 않은 것으로 나타났다. 즉, 설문조사 실시 간격에 관계없이 기준 시점과 비교 시점의 응답 비율이 통계적으로 유의한 차이를 보이지 않았다.

Table 1. Results of independence test

Range type	test-retest interval		
	2 weeks	4 weeks	2&4 weeks
Range1	0.755*	0.465	0.826
Range2	0.314	0.278	0.599
Range3	0.357	0.364	0.393
Range4	NA**	0.282	NA
Range5	0.314	0.319	0.364
Range6	0.389	0.243	0.465
Range7	0.343	0.675	0.627
Range8	0.314	0.555	0.364

\*: *p*-value

\*\* : not available due to perfect agreement between two paper-pencil tests

#### 3.2 Concordance rate

설문조사 연구를 실시한 시간 간격 및 피실험자별 응답의 일치율은 Table 2에 정리되어 있다. 2주 및 4주 간격 설문조사에서는 피실험자별로 두 번의 응답이 같은 경우, 2주 간격으로 2주 및 4주 연속 실시한 조사에서는 3회 응답이 모두 같은 경우를 일치하는 것으로 판정하였다. Table 2에는 본 연구에 참여한 피실험자의 가스렌지 형태별로 응답율이 가장 높은 버너-조종장치 연결(스테레오타이프)과 그 응답 비율이 병기되어 있다. 응답 일치율이 2주 간격으로 실시한 설문조사에서는 73.5~100.0%, 4주 간격으로 실시할 때는 67.1~97.6%, 그리고 2주 및 4주 연속 실시에서는 65.5~100.0%의 범위를 보였다. 2주 간격일 때의 일치율 평균은 92.3%(표준편차: 8.48), 4주 간격에서는 일치율 평

**Table 2.** Concordance rates by test-retest intervals

Range type	Concordance rate(%)			Stereotype			
	2 weeks	4 weeks	2&4 weeks	Proportion of response(%)			Stereotype
				2 weeks	4 weeks	2&4 weeks	
Range1	73.5	67.1	65.5	54.4	53.5	51.7	☆#△□
Range2	97.1	90.6	96.6	92.6	92.9	97.7	☆#△□
Range3	94.1	97.6	93.1	99.3	98.2	97.7	☆#△□
Range4	100.0	92.9	100.0	92.6	94.7	100.0	☆#△□
Range5	97.1	97.6	96.6	99.3	98.8	98.9	☆#△□
Range6	88.2	84.7	86.2	88.2	88.8	87.4	☆#□△
Range7	91.2	89.4	93.1	53.7	61.2	94.3	☆#△□
Range8	97.1	95.3	96.6	94.9	95.3	98.9	☆#△□

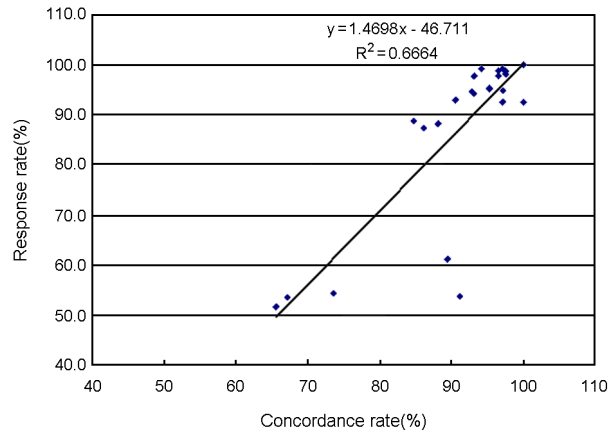
군 89.4%(표준편차: 10.02)를 보여, 조사 기간이 2주 길어지면서 일치율이 평균 약 2.9% 떨어졌다. 그 차이는 pairwise ttest에서 유의수준 5%에서 통계적으로 유의하였다( $p < 0.03$ ). 2주 및 4주 연속으로 실시한 설문조사 간 피실험자 응답 일치율은 평균 91.0%(표준편차: 10.02)로 나타났고, 2주 간격 및 4주 간격으로 실시한 설문조사의 일치율과 통계적으로 유의한 차이를 보이지 않았다( $p > 0.23$ ).

**3.3 Relationship between test-retest reliability and stereotype**

본 연구에서 실시한 3회의 설문조사 연구(2주, 4주 및 2, 4주 설문조사 실시)에서 스테레오타이프는 가스렌지 형태에 관계없이 모두 일치하였으며, 스테레오타이프 연결 형태에 대한 응답 비율도 비슷한 수준을 보였다(Table 2 참조). 가스렌지 형태별 스테레오타이프에 대한 응답률이 높을수록 설문조사 실시 간격별 피실험자 응답의 일치율도 높게 나타났다(Table 2, Figure 2). 이러한 관계는 각 가스렌지 형태별 피실험자 응답 일치율과 스테레오타이프 반응 비율 간의 상관계수가 0.82로 매우 강한 양의 상관관계를 보이는 데서도 볼 수 있다. 따라서 두 비율은 선형회귀식으로 잘 나타낼 수 있어( $R^2 = 0.67$ ) 한 변수로 다른 변수의 값을 추정할 수 있을 것으로 기대된다(Figure 2).

**4. Discussion and Conclusion**

본 연구에서는 4구 가스렌지 버너-조종장치 연결에 대한 스테레오타이프를 조사하는 설문조사의 신뢰성을 실험을 통



**Figure 2.** Concordance rates and proportion of responses for stereotype

하여 다루었다. 연구 결과 설문조사를 2주 간격, 4주 간격 혹은 2주 및 4주 간격으로 연속 실시하여도 처음 실시한 설문조사 결과와 응답 비율이 통계적으로 유의한 차이를 보이지 않았다. 즉, 설문조사 실시 간격이 피실험자 응답에 유의한 영향을 미치지 않았다. 따라서 4구 가스렌지 버너-조종장치 스테레오타이프를 조사하는 설문조사법의 재현성에는 문제가 없다고 할 수 있다.

가스렌지 형태별 버너-조종장치 스테레오타이프에 대한 응답 비율이 높을수록 설문조사의 재현성 즉, 피실험자 응답 일치율이 높게 나타났다(Table 2, Figure 2 참조). 어떤 스테레오타이프에 대한 피실험자의 응답 비율이 높다는 것은 스테레오타이프가 인간의 심성 모형(mental model) 혹은 직관에 가깝다는 것을 의미한다. 이로 인하여 일정 기간 간격을 두고 실시한 설문조사 간 피실험자의 응답 일치율도

자연스럽게 높게 나타났다고 해석할 수 있다.

각각 다른 피실험자가 참여한 2주 혹은 4주 간격으로 독립적으로 두 번 실시한 설문조사 응답 일치율과 같은 피실험자가 참여한 2주 및 4주 연속으로 실시한 응답 일치율이 통계적으로 유의한 차이를 보이지 않았다(Table 2 참조). 이는 설문조사를 연속적으로 세 번을 실시하여도 반복 실시로 인한 학습효과와 영향이 크지 않음을 의미한다. 또한, 두 번째 또는 세 번째 설문조사 시에 전 설문조사의 응답 내용을 떠올리려 하지 말고 현재의 직관으로 답하라고 한 주의사항이 피실험자에게 제대로 전달된 것으로도 해석할 수 있다.

본 연구에서 4구 가스렌지 버너-조종장치 스테레오타이프 조사를 위한 설문조사를 2주 혹은 4주 간격으로 실시할 때 높은 재현성을 보였다. 이는 1~4주 간격으로 조사를 실시하는 것은 수용할만하다고 한 전 연구(Coluci et al., 2009)의 주장과 일치하고, 스테레오타이프 조사를 위한 설문조사도 4주 이내에 실시할 경우 높은 재현성을 얻을 수 있음을 의미한다.

## References

- Chapanis, A. and Lindenbaum, L.E., A reaction-time study of four control-display linkage, *Human Factors*, 1, 1-7, 1959.
- Coluci, M.Z.O., Alexandre, N.M.C. and Rosecrance, J., Reliability and validity of an ergonomics-related Job Factors Questionnaire, *International Journal of Industrial Ergonomics*, 39, 995-1001, 2009.
- d'Errico, A., Gore, R.G., Gold, J.E., Park, J-S. and Punnett, L., Medium- and long-term reproducibility of self-reported exposure to physical ergonomics factors at work, *Applied Ergonomics*, 38, 167-175, 2007.
- Hlapern, M., Nordin, H.M., Goldsheyder, D. and Crane, M., The test-retest reliability of a new occupational risk factor questionnaire for outcome studies of low back pain, *Applied Ergonomics*, 32, 39-46, 2001.
- Hoffmann, E.R., Do paper-and-pencil tests give an accurate measure of stereotype strength?: A review of available data, *International Journal of Industrial Ergonomics*, 39, 904-912, 2009.
- Hsu, S-H. and Peng, Y., Control/display relationship of the four-burner stove: A reexamination, *Human Factors*, 35(4), 745-749, 1993.
- Ijmker, S., Mikkers, J., Blatter, B.M., van der Beek, A.J., van Mechelen, W. and Bongers, P.M., Test-retest reliability and concurrent validity of a web-based questionnaire measuring workstation and individual correlates of work postures during computer work, *Applied Ergonomics*, 39, 685-696, 2008.
- Kee, D., Stereotype for control-display relationship of four-burner range for Korean, *Journal of the Ergonomics Society of Korea*, 30(3), 389-394, 2011a.
- Kee, D., Stereotype and effective cues for burner-control relationship of four-stove range, *Journal of the Korean Institute of Industrial Engineers*, 37(2), 118-123, 2011b.
- Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N. and Yamamura, T., Physiological response to workload change. A test/retest examination, *Applied Ergonomics*, 40, 987-996, 2009.
- Shinar, D. and Acton, M.B., Control-display relationships on the four-burner range: Population stereotypes versus standard, *Human Factors*, 20, 13-17, 1978.
- Wu, S-P., Further studies on the spatial compatibility of four control-display linkages, *International Journal of Industrial Ergonomics*, 19, 353-360, 1997.

## Author listings

**Doyung Kee:** dhkee@kmu.ac.kr

**Highest degree:** Ph.D., Department of Industrial Engineering, POSTECH

**Position title:** Professor, Department of Industrial and Management Engineering, Keimyung University

**Areas of interest:** Population stereotype, product design, posture classification scheme, industrial safety, musculoskeletal disorders

Date Received : 2013-05-01

Date Revised : 2013-05-16

Date Accepted : 2013-05-17