

# 문헌정보학 연구에 있어서 타당도 및 신뢰도에 관한 연구

## A Study on the Validity and Reliability in the Research of Library and Information Science

김 성 희(Seong-Hee Kim)\*

### 목 차

- |                         |   |
|-------------------------|---|
| 1. 서 론                  | 2. 2. 3 반분검사 신뢰도                        |
| 2. 타당도 및 신뢰도에 대한 이론적 배경 | 3. 타당도 및 신뢰도에 영향을 미치는<br>요인분석           |
| 2. 1 타당도 유형 및 추정방법      | 3. 1 타당도에 영향을 미치는 요인분석                  |
| 2. 1. 1 내용타당도           | 3. 2 신뢰도에 영향을 미치는 요인분석                  |
| 2. 1. 2 구인타당도           | 3. 3 타당도와 신뢰도의 관계                       |
| 2. 1. 3 준거타당도           | 4. 문헌정보학 연구에서 타당도 및 신뢰도<br>적용시 고려해야할 요소 |
| 2. 2 신뢰도 유형 및 추정방법      | 5. 결 론                                  |
| 2. 2. 1 재검사 신뢰도         |   |
| 2. 2. 2 동형검사 신뢰도        |   |

### 초 록

본 논문에서는 문헌정보학의 과학적인 연구결과를 얻는데 있어서 중요한 요소인 타당도와 신뢰도에 대해서 검토하고 있다. 먼저 연구의 타당도와 신뢰도에 대한 개요 및 유형을 살펴본 뒤 질문지 또는 다양한 검사문항을 작성하는데 있어서 타당도와 신뢰도에 영향을 미치는 요인들을 과거의 선행연구를 통해 분석하고 있다. 이런 연구결과는 미래의 문헌정보학과 관련된 문제들을 과학적으로 해결하기 위한 기초자료로써 활용할 수 있을 것이다.

### ABSTRACTS

This study describes validity and reliability, which are important factors to obtain scientific results in the study of library and information science. Specifically, this paper starts by exploring the definition and types of validity and reliability. It then analyzes the factors which affect to validity and reliability in developing test-items through the previous studies. It is proposed that validity and reliability be regarded as basic and important components to solve problems related to library and information science in the future scientifically.

\* 동덕여자대학교 정보대학 정보학부 조교수  
접수일자 2000년 6월 2일

## 1. 서론

하나의 학문분야가 독립된 과학으로서 성립하려면 두 가지 조건을 충족시킬 수 있어야 한다. 첫째는 그 학문 고유의 연구대상 즉 연구할 문제들이 있어야 하고, 둘째로는 그 연구대상을 탐구해서 그에 대한 이론을 세울 수 있는 과학적 방법이 갖추어져야 한다(Vnacke et al., 1964). 이러한 조건이 갖추어졌을 때 연구문제들을 과학적 방법을 통하여 해결하고 그 결과들을 한데 묶어 하나의 체계화된 이론을 형성하게 되어 하나의 독립된 과학이 성립된다고 할 수 있다. 한편 모든 학문의 연구목적은 주어진 현상에 관련된 변인들과 이들 변인들간의 관계를 ①기술(describe)하고 ②설명(explain)해서 이에 대한 보편 타당한 지식 체계 즉 이론을 정립함으로써 그 주어진 현상을 미리 ③예측(predict)하고 나아가서는 그것을 ④통제(control)하는데 있다.

현재 문헌정보학은 전통적인 '도서관학'이 '정보학'과 접목됨으로써 미래지향적인 학문으로 발전하고 있다. 실제로 문헌정보학의 최근발전추세를 보면 커뮤니케이션, 컴퓨터공학 및 통신, 인공지능, 도서관학, 정보학, 언어학, 심리학, 응용통계학, 경영학 등 관련 학문과의 접목이 이루어지면서 문헌정보학을 기반으로 한 학문적 상부구조가 발전적으로 형성되어 왔다. 또한, 문헌정보학이 하나의 독립된 학문 분야로 정립되고, 학문적으로 성장해오면서 다양한 연구문제들을 해결해 왔으며 연구방법도 다양한 방법을 적극 수용하여 다각적인 연구가 수행되고 있다. 이와 같이 논문의 양의 증가함에 따라 그 논문의 내용의 질을 평가할

필요가 있다. 아무리 훌륭한 글이라도 빈약하게 설계되거나 집행되어진 연구는 그 질이 낮아질 수밖에 없다. 사실 이러한 결점이 논문이 학술지에 게재여부를 심사하는데 중요한 게재불가요인이 될 수 있다. 따라서, 연구논문의 질적인 면을 평가하고 이 연구가 중요하고 그리고 발표하기에 결점이 없는지를 스스로 검토해보아야 할 것이다. 만일, 이 논문이 다른 연구자의 것이라면 과연 내가 이것을 읽을 것인가? 이것이 나의 연구에 영향을 줄 것인가? 대부분의 연구자 서랍에는 이런 기준에 미달한 한 두 개 이상의 연구가 들어 있을 것이다. 얼마나 잘 쓰여졌던 간에 연구방법이 미흡하면 연구는 받아들여질 수 없다. 즉, 연구결과는 충분한 객관적 근거와 논리와 설득력을 가지고 있어야 하고 객관적 타당도와 신뢰도를 가지고 있어야 한다. 따라서, 연구결과를 평가하는데 가장 중요한 요소중의 하나가 타당도와 신뢰도라고 할 수 있다. 일반적으로 말해서, 연구결과가 사실일 때 타당하다고 하고 그 결과들을 반복할 수 있을 때 신뢰도가 있다고 말할 수 있다. 따라서, 타당도가 높다고 하는 것은 연구결과가 사실이라는 의미이고 신뢰도가 높다고 하는 것은 연구결과가 여러 번 반복해서 시행해도 같은 결과를 얻음으로써 일관성이 있다는 것이다. 따라서, 신뢰도와 타당도는 연구결과의 평가에서 실질적으로 매우 중요한 요소이다. 즉, 과학적인 연구결과를 얻기 위해서 연구자는 연구설계에서 그 결론들이 사실이고 반복할 수 있는지를 물어야 하며, 측정 면에서 그 관찰이나 점수가 정확하고 반복할 수 있는지 확인해야 한다. 그렇지만 아직까지 문헌정보학 관련 학술 논문들

중에는 타당도와 신뢰도에 관한 검증은 거의 이루어져 오지 못한 실정이다. 일반적으로 타당도와 신뢰도 검증 없이 연구결과를 도출한다는 것은 연구결과가 일반화되기 어렵고 과학적이고 객관적인 결론을 도출할 수 없다. 따라서, 본 논문에서는 문헌정보학분야의 다양한 연구주제를 선정해서 과학적인 연구결과를 얻는데 있어서 중요한 요소인 타당도와 신뢰도에 대해서 검토해보고자 한다. 먼저 연구의 타당도와 신뢰도에 대한 개요를 살펴본 뒤 질문지 또는 다양한 검사문항을 작성하는데 있어서 타당도와 신뢰도에 영향을 미치는 요인들을 과거의 선행연구를 통해 분석해보고자 한다. 이런 연구결과는 미래의 문헌정보학과 관련된 문제들을 과학적으로 해결하기 위한 과학적 연구의 기초자료로써 활용할 수 있을 것이다.

## 2. 타당도 및 신뢰도에 대한 이론적 배경

### 2.1 타당도유형 및 추정방법

타당도란 측정하고자하는 것을 실제로 측정하고 있는가? 와 관련된 것으로 올바른 측정이 되기 위해서는 측정하고자 하는 것을 실제로 측정되어야만 한다.

타당도에 관한 정의는 학자에 따라 다양하게 정의되고 있다. 먼저, Anastasi(1954)가 타당도를 안면타당도(face validity), 내용타당도(content validity), 요인타당도(factorial validity), 경험타당도(empirical validity)로 구분하였으나, 제2판

(1961)에서 타당도를 정리하여 내용타당도(content validity), 예언타당도(prediction validity), 공인타당도(concurrent validity), 구인타당도(construct validity)로 구분하였다. Anastasi의 제3판(1968), 제4판(1976), 제5판(1982), 제6판(1988)을 거치면서 타당도를 3종류로 구분하였다. 즉, 내용관련타당성(content-related validation), 준거관련타당성(criterion-related validation), 구인관련타당성(construct-related validation)으로 구분하였다.

Cronbach(1949)는 타당도를 논리적 타당도(logical validity), 경험타당도(empirical validity)로 구분하였으나 그의 저서 2판(1960)에서 Anastasi의 제2판(1961)과 같이 타당도를 네 종류로 구분하였고 제3판(1970)과 제4판(1984), 제5판(1990)에서 세 종류의 타당도로 분류하여 내용관련 타당도(content-related inquiry), 준거관련 타당도(criterion-related inquiry), 구인관련 타당도(construct-related inquiry)로 명하고 있다. 다음은 이상의 다양한 타당도 유형들을 크게 내용타당도, 구인타당도, 준거타당도로 구분해서 설명하고 있다.

#### 2.1.1 내용타당도

내용타당도(content validity)는 측정도구가 갖고 있는 내용, 즉 실제, 소재, 논제의 대표성(representativeness)을 의미한다. 즉 측정내용이 측정하려는 속성내용의 전집을 대표하는 것인가? 하는 것이다. 따라서 내용타당도는 검사의 문항 측정을 위하여 규정된 내용영역이나 혹은 전체를 얼마나 잘 대표하느냐의 정도를 나타낸다고 할 수 있다. 내용타당도를 검증하는

방법은 검사내용전문가에 의하여 내용타당도가 있다 혹은 없지로 표현한다. 따라서, 내용타당도는 전문가의 판단에 의존하므로 객관적 자료를 사용하지 않으며 타당도에 대하여 수치로 나타내지 않는다. 단점으로는 전문가에 따라 내용타당도에 대한 각기 다른 검증결과를 얻을 수 있다. 따라서, 전문가의 전문적 판단이 측정 내용의 전집, 내용선택, 문항유형선택, 점수화등의 의사결정을 하는데 중요한 역할을 한다.

### 2. 1. 2 구인타당도

구인이란 심리적 특성이나 행동양상을 설명하기 위하여 존재를 가정하는 심리적 요인을 말한다. 예를 들면, 사서의 창의성 검사의 예에서 민감성, 이행성, 도전성등을 구인이라 할 수 있다. 지능검사에서 어휘력, 수리력, 추리력, 공간력, 지각력, 기억력, 언어유창성이 구인이 될 수 있다. 도서관장의 지도력에 대한 연구에서 지도력의 구인으로 칭찬을 하는 행위, 어떤 행위에 대한 정당성 설명, 의문제기, 목적완수, 계획집행등이 구인으로 고려될 수 있다. 따라서, 구인 타당도(construct validity)란 조작적으로 정의되지 아니한 인간의 심리적 특성이나 성질을 심리적 구인으로 분석하여 조작적 정의를 부여한 후, 검사점수가 조작적 정의에서 규명한 심리적 구인들을 제대로 측정하였는가를 검증하는 방법이다. 예를 들면, 창의성을 측정할 때 민감성, 이행성, 도전성등의 구인으로 구성되어 있다는 조작적 정의에 근거하여 검사문항을 제작, 실시하여 그 검사문항이 이와 같은 구인들을 측정하고 있다면 그 검사문항은 구인타당도를 지닌다고 할 수 있다. 그러

나 만약 검사 결과가 조작적으로 규정한 어떤 심리적 특성의 구인들을 제대로 측정하고 있지 못하거나 다른 구인들을 측정한다면 이는 구인 타당도가 결여되어 있는 것이다. 구인타당도를 추정하는 방법은 다음과 같다.

첫째, 측정하고자 하는 심리적 특성을 구성하는 구인, 즉 요소들을 이론적, 경험적 배경에 의하여 밝힌다. 즉 심리적 특성에 대한 조작적 정의를 내린다.

둘째, 구인을 측정할 수 있는 문항을 제작한다.

셋째, 측정대상에게 검사를 실시하여 응답 자료를 얻는다.

넷째, 응답자료를 분석하여 검사가 측정하고자하는 구인들을 제대로 측정하였는지를 밝힌다.

다섯째, 심리적 특성을 규명하는 조작적 정의에 관계되는 구인과 관계가 없는 문항을 삭제한다.

구인타당도를 검증하는 통계적 방법으로는 상관계수법, 실험설계, 요인분석등을 들 수 있다. 이들 기법 중에서 상관계수법을 예로 들면 사서의 창의성은 민감성, 이해성, 도전성, 개방성, 자발성, 자신감, 그리고 암기능력으로 구성되어 있다고 조작적 정의를 내렸다면 각 구인에 의한 점수와 창의성 총점과의 상관계수를 구할 수 있다. 각기 10문항으로 각 구인을 측정하고 70문항 점수를 각 구인별로 점수를 내어 계산한 후 총점과 상관계수를 추정하고 각 구인들의 점수간의 상관계수를 추정한 결과 <표 1>과 같다.

위의 표에서 창의성을 구성하는 7개 구인 중에 암기력을 제외한 구인들은 창의성 총점

〈표 1〉 창의성을 구성하는 구인과 총점간의 상관계수

	민감성	이해성	도전성	개방성	자발성	자신감	암기력
이해성	.7						
도전성	.8	.7					
개방성	.9	.8	.9				
자발성	.8	.8	.8	.8			
자신감	.7	.9	.9	.9	.8		
암기력	.2	.3	.3	.1	.3	.2	
총점	.8	.9	.9	.8	.7	.9	.2

과 상관계수 .7이상으로 높는데 비하여 암기력은 창의성 총점과 상관계수가 .2로 거의 없다고 할 수 있다. 또한 각 구인들간의 상관계수가 높는데 비하여 암기력은 다른 여섯 개의 구인들과의 상관계수가 낮다. 그러므로 암기력은 창의성을 나타내는 구인이 될 수 없다고 판단하게 된다.

구인타당도는 응답자료에 의하여 계량적방법에 의하여 검증되므로 과학적이고 객관적이라 할 수 있다. 또한 모르는 심리적 특성에 부여한 조작적 정의의 타당성을 밝혀 주므로 많은 연구의 기초가 될 수 있다. 단점으로는 안정적인 상관계수를 얻기 위하여 많은 연구대상이 필요하다.

### 2. 1. 3 준거타당도

준거타당도(criterion validity)는 측정도구에 의한 점수와 어떤 준거와의 상관계수에 의하여 측정도구의 타당도를 검증하는 방법으로 공인타당도(concurrent validity)와 예측타당도(predictive validity)가 있다.

#### (1) 공인타당도

공인타당도는 새로운 측정도구를 제작하였

을 때 새로 제작한 측정도구의 타당성을 검증하기 위해 기존에 타당성을 보장받고 있는 검사와의 유사성에 의해 타당성을 검증하는 방법이다. 공인타당도는 새로 제작한 검사문항에 의한 점수와 준거점수로 타당성을 인정받고 있는 검사의 점수와 상관계수에 의하여 검증되므로 계량화될 수 있다. 공인타당도 추정절차는 다음과 같다.

첫째, 피험자 집단에게 새로 제작된 측정도구를 이용해서 검사를 실시한다.

둘째, 동일집단에게 동일한 상황에서 타당성을 인정받고 있는 검사를 실시한다.

셋째, 이들 두점수간의 상관계수를 추정한다.

예를 들면, 어떤 사서가 인터넷 정보검색사 능력을 간단히 측정하는 간편한 검사도구를 개발하였다. 5문항으로 구성된 이 검사의 공인타당도를 검증하기 위하여 5명의 학생에게 검사를 실시하고, 이들에게 10문항으로 구성된 표준화 인터넷 검색능력 검사를 실시한 결과 두 검사 점수간의 상관계수가 .9라면 공인타당도가 매우 높은 편이라고 해석할 수 있다.

공인타당도는 계량화되어 타당도에 대한 객관적인 정보를 제공할 수 있으며 타당도의 정도를 나타낼 수 있는 장점을 갖는다. 단점으로

는 기존의 타당성을 입증 받고 있는 검사가 없을 경우 공인타당도를 추정할 수 없다.

## (2) 예측타당도

예측타당도(predictive validity)는 주어진 측정의 결과가 장래의 어떤 효과나 행동을 얼마나 잘 예언해 낼 수 있는지의 정도를 실증적 자료를 기초로 통계적으로 밝히는 방법이다. 예를 들어, 대학 신입생 집단에 지능검사를 실시하여 점수를 구한 다음, 이들 신입생이 1년동안에 받은 학업평점의 평균과 비교하여 지능지수와 학업평점간의 상관관계가 높으면 지능검사는 예측적 타당도가 높다고 할 수 있다. 일반적으로 적성검사나 심리검사가 예측타당도를 중시하는 경향이 있다. 예측타당도도 역시 검사점수와 미래의 행동과의 관계에 의하여 추정되므로 계량화되는 특징이 있으며 추정방법도 공인타당도와 같이 상관관계수에 의하며 그 절차는 다음과 같다.

첫째, 피험자 집단에게 새로 제작한 검사를 실시한다.

둘째, 일정기간 후 검사한 내용과 관계가 있는 피험자들의 행위를 측정한다.

셋째, 검사점수와 미래 행위의 측정치와 상관정도를 추정한다.

예를 들면, 사서의 인터넷 검색 적성시험을 실시한 결과와 인터넷 검색교육을 마친 후 인터넷 검색사 시험성적결과 이 둘의 상관관계수가 .7이 나왔을 경우 예측타당도가 있다고 설명할 수 있다.

주의해야할 사항은 미래 행위에 대한 측정이다. 미래의 행위에 대한 측정은 이전에 만들었던 검사로 다시 측정하는 것이 아니라 검사

에서 특정한 내용과 관련된 행동이어야 한다. 예측타당도의 장점으로는 검사도구가 미래의 행위를 예언하여 주기 때문에 예측타당도가 높으면 선발, 채용, 배치 등의 목적을 위하여 검사를 사용할 수 있다. 단점으로는 동시 측정이 불가능하므로 검사의 타당성을 검증하기 위하여 시간적 여유가 있어야 하며 일정시간 뒤 측정할 경우 여러 가지 요인이 개입될 수 있다는 것이다.

## 2. 2 신뢰도 유형 및 추정방법

신뢰도(reliability)란 둘 이상의 관찰치가 각기 동일한 관찰에서 동일한 결과를 얻을 수 있는 정도로, 타당성과 더불어 과학적인 연구결과를 얻는데 가장 중요한 요소이다. 어떤 연구결과가 신뢰롭다는 것은 그 연구자료가 안정되고(stability), 일관성(consistency)이 있으며, 정확(accuracy)하고 예측가능성(predictability)이 있다는 것이다(Kerlinger 1984). 신뢰도는 대체로 3가지 뜻을 함축하고 있다.

첫째 동일한 대상들의 집합을 동일하거나 비교 가능한 측정도구로 반복해서 측정한다면, 동일하거나 서로 비슷한 결과를 얻을 수 있겠는가?하는 것으로 안정성(stability), 예측가능성(predictability)과 관련이 된다. 둘째로는 측정도구로 부터 얻은 측정치가 측정된 속성의 "실제" 측정치인가?와 관련된 것으로 정확성(accuracy)과 결부된다. 마지막으로 신뢰도는 오차와 관련된 것으로 신뢰도를 높이기 위해서는 오차를 줄여야 한다.

신뢰도의 유형으로는 측정도구의 안정성을 측정하는 재검사 신뢰도, 유사성을 측정하는

동형검사 신뢰도, 그리고 분산의 비율개념에 의한 문항간 내적 일관성 신뢰도가 있다.

신뢰도추정방법으로는 Karl Pearson의 단순적률 상관계수 추정공식에 의한 재검사 신뢰도, 동형검사 신뢰도가 있으며 Spearman과 Brown의 반분검사 신뢰도 등이 있다.

### 2. 2. 1 재검사 신뢰도

재검사 신뢰도(test-retest reliability)는 동일한 검사를 동일한 피험자 집단에 일정 시간 간격을 두고 두 번 실시하여 얻은 두 검사점수의 상관계수에 의하여 신뢰도를 검증하는 방법이다. Karl Pearson의 단순적률 상관계수 추정공식에 의하여 신뢰도 계수가 산출되며, 측정도구의 안정성을 측정하게 된다. 예를 들면, 10문항으로 제작된 사서의 인터넷 활용능력 시험의 재검사 신뢰도를 추정하기 위하여 6명의 학생에게 두 번의 검사를 실시하여 얻은 점수결과 검사와 재검사에서 얻은 점수의 상관계수가 .9라고 했을 경우 인터넷 활용능력 검사의 재검사 신뢰도는 매우 높다고 할 수 있다. 재검사 신뢰도를 위한 시험간격은 일반적으로 피험자의 기억이 소멸되는 시간으로 2주에서 4주로 설정하는 것이 바람직하다. 재검사 신뢰도는 추정방법이 간단하나 시험간격설정의 문제와 동일검사로 인한 기억이나 연습효과가 작용될 가능성이 많다는 것이다.

### 2. 2. 2 동형검사 신뢰도

동형검사 신뢰도(parallel-form reliability)는 측정도구를 만들 때 하나만 만들지 않고 비슷한 두 개를 만들어 동시에 또는 일정한 시일을 두고 같은 집단에 실시하여 그들 두

결과 사이의 상관계수를 가지고 Spearman-Brown공식에 의하여 신뢰도를 추정하는 방법이다. 동형검사의 기본 조건은 한 검사에 의한 진점수와 다른 동형검사에 의한 진점수가 같아야 하고, 한 검사에 의한 오차점수의 분산과 동형검사에 의한 오차점수 분산이 같아야 한다. 또한 동형검사 추정을 위해 검사를 제작할 때 동일내용을 측정하여야 하며, 동일한 형태의 문항과 문항 수, 그리고 동일한 문항난이도와 문항 변별도가 지켜져야 한다는 것이다.

추정방법으로는 두 개의 동형검사가 제작되어 동일피험자에게 검사가 두 번 시행되어야 한다. 예를 들면, 10문항으로 제작된 이용자 만족도 검사와 동형검사를 5명의 피험자에게 실시한 결과 상관계수가 .8이 나왔다고 하면 동형검사 신뢰도가 높다고 할 수 있다.

동형검사 신뢰도는 두 개의 동형검사를 동일한 집단에 동시에 시행하므로 시험간격이 문제되지 않고 신뢰도 계수추정이 쉽다. 반면 이 방법의 단점은 완전히 똑같은 형태의 두개의 설문내용을 만들기 어렵다는 것이다. 또한, 재검사 신뢰도처럼 검사를 두 번 시행하는데 따른 문제점이 있다. 즉, 동일한 검사환경, 피험자의 동일한 검사동기와 검사태도를 만들기 어렵다.

### 2. 2. 3 반분검사 신뢰도

재검사 신뢰도와 동형검사 신뢰도는 동일 피험자에게 검사를 두 번 실시하는 번거로움이 따르며, 시험간격과 검사의 동형성 정도에 따라 신뢰도 계수가 변화되는 문제점을 갖고 있다. 이에 비하여 반분검사(split-half reliability)는 검사문항들을 만든 다음 이들 문항들을 반으

로 나누어서 두개의 측정처럼 따로 채점하여 그 사이의 상관관계를 산출하여 Spearman-Brown공식에 의하여 신뢰도를 검토하는 방법이다.

반분검사 신뢰도를 추정할 때 검사를 양분한 두 부분검사점수의 상관계수를 그대로 사용하면 신뢰도가 과소 추정된다. 만약 20문항으로 제작된 측정도구를 10문항씩 양분하여 두 부분 검사 점수간의 상관계수를 추정하였다면 이는 20문항의 검사 신뢰도가 아니라 10문항으로 구성된 동형검사가 되어 신뢰도 계수가 감소하게 된다.

반분검사 신뢰도 추정이 검사를 반분하는 방법에는 짝수문항과 홀수문항으로 반분하는 기우법과 전체검사를 문항순서에 따라 전과 후로 나누는 전후법, 무작위로 분할하는 단순 무작위법, 문항특성에 의하여 반분하는 방법 등이 있다. 문항특성에 의한 반분법은 문항난이도와 문항변별도에 의존한다. 문항 특성에 의하여 검사를 반분하면 동일한 난이도와 동일한 점수를 갖는 두 부분 검사를 만들 수 있다. 예를 들면, 6문항으로 제작된 검사를 5명의 피험자에게 실시하여 얻은 검사의 문항점수와 총점이 있을 경우 홀수와 짝수문항으로 양분하여 반분검사 신뢰도를 추정하였다면 계수가 .9라 하였을 경우 반분검사 신뢰도는 매우 높다고 할 수 있다.

반분검사는 재검사나 동형검사 처럼 두 번 시행하지 않으므로 시험간격이나 동형검사제작에 문제가 되지 않는다는 장점이 있다. 그러나 검사를 양분하는 방법에 따라 신뢰도 계수가 달리 추정되므로 신뢰도를 높게 추정하기 위해서는 문항특성에 의하여 검사를 반분하는

것이 바람직하다.

### 3. 타당도 및 신뢰도에 영향을 미치는 요인분석

#### 3.1 타당도에 영향을 미치는 요인분석

검사문항을 작성할 경우 타당도에 미치는 요인은 크게 내용과 문항의 표현으로 구분할 수 있다. 먼저 검사문항 및 내용과 관련하여 여러 가지 검사문항들은 관찰하려는 모집단들로부터 얻어진 표본을 대상으로 실시한다 (Cronbach 1957). 예를 들면, 도서관 이용 만족도와 관련된 어떤 검사항목을 될 수 있는 한 많은 수의 적절한 문항들로부터 표집된 제한된 문항들만 다룬다. 그러므로 특정조사를 위한 질문 문항의 내용의 적절성은 연구결과들을 예측하는데 매우 중요하다(Cronbach 1957).

한편, 내용의 적절성은 특정모집단을 정의하고 또 질문지에 정확하게 표현함으로써 얻어진다. 특정모집단을 표집된 실제검사에 얼마나 정확하게 나타내느냐의 여부가 검사의 타당도를 크게 변화시킬 수 있을 것이다. 즉, 어떤 검사문항들이 실제로는 검사자가 측정하려는 내용들을 포함하고 있지만 그 내용을 어떻게 진술하느냐에 따라 피험자들에게 그 내용을 전달하는 효과가 달라질 수 있다. 따라서, 검사문항은 모든 응답자들이 똑같이 해석할 수 있도록 진술해야 한다. 이러한 점을 충족시키기 위해서 문항의 진술은 보다 구체적이고 특수한 형태를 취해야 한다. 다음은 검사



문항 표현과 관련해서 타당도에 영향을 미치는 요인들이다.

### 3. 1. 1 빈도수식어와 타당도

빈도수식어강도란 문항이 표현하려는 어떤 상황의 발생빈도를 나타내는 낱말이다. 사용될 수식어들은 가끔, 때때로(낮은 강도), 일반적으로, 종종(중간강도), 그리고 자주, 항상(높은 강도)등이다. 질문지 등을 포함해서 검사 문항의 진술에 사용되는 빈도수식어들은 특정 문항이 표현하려는 어떤 상황을 구체적으로 또는 모호하게 만들 수도 있다. 예를 들면, “연구하고 싶을 때가 있다”라는 진술과 “연구하고 싶을 때가 자주 있다”라는 진술을 생각할 경우 이들간의 진술은 응답자들에게 서로 다른 느낌을 갖게 할 수도 있다. 보기의 진술에서 표현된 “자주”라는 수식어가 전체문항의 초점을 변화시켜 그 문항에 대한 응답자들의 반응에 영향을 미칠 수도 있다. 즉 어떤 한 사람은 관심 있는 행동에 반응하는 한편, 다른 한 사람은 특정한 빈도수식에 반응할 수도 있다는 뜻이다.

Simpson(1944)은 빈도수식어들의 양적 의미 차이(quantitative meaning gap)가 개인들

간에 존재함을 밝혔다. 그는 학생들에게 “대개 어떤 일을 했다”는 말의 양적 의미를 물었다. 25%의 학생들이 시간의 90% 정도의 빈도(frequency)로 반응하는 것이라고 답했고, 나머지 25%의 학생들은 시간의 70%이하의 빈도로 반응하는 것이라고 답하였다. 다른 빈도수식어들에 대한 반응자들의 양적 해석이 <표 2>에 제시되어 있다. 동일한 행동을 가진 피험자들이 같은 빈도수식어에 대한 상당히 다른 양적 해석을 한다는 사실이 드러났다.

Altermeyer(1948)은 문항의 진술에 사용된 이러한 빈도수식어들의 의미차이가 데이터에 영향을 미친다고 하였다.

이상에서와 같이 문항에 사용될 빈도수식어들의 강도만 변화 시켜도 문항의 반응분포를 변화시킬 수 있음을 시사하고 있다. 일반적으로 질문문항 작성시 중간정도의 신뢰도와 타당도가 가장 높다고 나타났다(Deaton 1980). 특히, 반응자들은 문항의 모호성이 높은 ‘높은 강도’의 수식어들이 삽입된 문항들에 대해서는 특정 평균치를 향해 반응하는 경향이 있다. 즉, 응답자들의 반응이 문항 반응 평균을 중심으로 밀집된 형태를 보인다는 뜻이다. 따라서, 문항 제작시 낮은 강도의 빈도수식어를 사용

<표 2> 질문지에서 사용되는 빈도수식어에 대한 양적 의미 범위

빈도수식어	중앙치(%)	응답자중 중간 50%의 대답의 범위(%)
usually	85	70 - 90
often	78	65 - 85
frequently	73	40 - 80
sometimes	20	13 - 35
occasionally	20	10 - 33
seldom	10	6 - 18
rarely	5	3 - 10

하거나, 빈도수식어를 사용하지 않는 것이 바람직할 것으로 보인다. 특히, “일반적으로” 등과 같은 심리학적 연속체 상에서의 위치가 모호한 빈도수식어들의 사용은 피하는 것이 좋을 것이다.

### 3. 1. 2 문항의 길이와 타당도

보통 12어절 이상으로 진술된 문항을 긴 문항으로, 12어절 이하로 진술된 문항을 짧은 문항으로 구분한다. 정희태(1987)의 연구에 의하면 먼저 문항반응평균들이 분석단위로 사용되었을 때 통계적으로 유의한 차가 짧고 긴 문항들간에 나타났다. 그리고 응답자들은 짧은 길이의 문항들에 대해 보다 확실한 해석을 하게 되고 그 결과로 짧은 문항들에 대해서는 각 척도치들에 대해 골고루 반응하는 경향이 있다. 반면에 문항의 길이가 긴 문항에는 어절 수가 길어짐에 따라 해석의 모호성을 증대시켜 특정 평균치로 향하여 반응하는 경향이 있었다. 이런 결과는 Edwards(1948) 등은 “문항의 진술은 짧아야 하며, 20단어 내외로 해야 한다”고 하는 주장과 일치하고 있다. 따라서, 문항의 진술을 짧게 하는 것이 바람직할 것으로 보인다.

### 3. 1. 3 문항의 진술방향과 타당도

문항의 진술방향은 정적방향으로 진술한 문항을 정적문항, 부적방향으로 진술한 문항을 부적문항으로 구분한다. 기존의 연구에 따르면 문항의 진술방향에 따라 문항반응평균들이 차이를 보였다. 반응자들은 정적방향으로 진술된 문항들에 대해서는 각 척도치들에 대해 골고루 반응하는 경향이 있는 반면 부적방향으로

진술된 문항들에 대해서는 어떤 특정 평균치를 향하거나 극단 값으로(1이나 5) 반응하는 경향이 있었다. Borg와 Gall(1983)은 부적 문항들은 많은 응답자들이 오독(misread)할 수 있으므로 피해야 한다고 주장했다. 즉 응답자들은 부적인 단어를 무시(overlook)하고 그의 견해와 반대되는 답을 할 위험이 있다는 것이다. 그리고 Whipple(1957)과 Cramer(1951)의 연구 결과가 개인들의 반응들이 진술의 부적이거나 정적인 어구에 의해 영향을 받는다고 하였다.

Whipple(1957)은 정적(positive)이거나 부적(negative)인 방향의 구절을 가진 진위형 문항들으로써 연구한 결과, 반응자들이 정적인 방향으로 진술된 문항들에 대해서는 진이라고 답하는 경향이 있었음을 발견하였다. Cramer(1951)도 시사질문지(current event questionnaire)에 대해 비슷한 방법을 적용하였는데, 어떤 내용을 부적으로 진술했을 때는 반응자들이 거부하는 경향이 있었고, 같은 내용을 정적으로 진술했을 때에는 동의하는 경향이 있었음을 발견하였다.

## 3. 2 신뢰도에 영향을 주는 요인분석

신뢰도에 영향을 주는 요인들로는 검사에 관련된 요인, 피험자에 관련된 요인, 검사 시행에 관련된 요인들을 들 수 있다. 먼저, 피험자와 관련된 요인들로 피험자 집단의 이질성, 피험자들의 검사에 대한 지혜, 피험자들의 흥미와 동기를 들 수 있다. 피험자 집단이 이질적일 때 신뢰도는 증가한다. 또한 검사에 대한 흥미가 높고 검사선택 동기가 높으면 신뢰도는 증가하나 흥미를 상실한 상태에서 검사를

실시하면 응답의 일관성을 결여하여 신뢰도가 감소하게 된다. 검사 시행에 관련된 요인들로 는 시간제한, 시험 환경 등이 검사의 신뢰도를 증가 시키기도하고 감소시키기도 하는 요인들 로 알려져 있다.

일반적으로 태도검사를 2세이하의 아동들 에게 실시할 때 신뢰도 정보는 무익하며, 5세 아동들에게는 개인검사를 실시할 때 신뢰도를 검증할 수 있다. 집단검사는 최소 10세 이상의 아동들에게 실시하여야 검증할 수 있다. 다음 은 검사에 관련된 요인들이다.

### 3. 2. 1 문항수와 신뢰도

일반적으로 문항수는 신뢰도에 영향을 미친 다. 즉, 적은수의 문항으로 인간이 지니고 있 는 속성을 측정할 때보다 많은 수의 문항으로 측정할 때 측정의 오차를 줄일 수 있다. 물론 증가되는 문항은 양질의 문항으로 동일 특성 을 측정하고, 문항난이도가 유사하며, 문항변 별도가 높아야 한다. 양질의 문항수를 증가한 다 하여 신뢰도가 계속 선형적으로 증가하는 것이 아니라 곡선형적으로 증가한다. 문항수 를 늘려서 검사를 길게 하였을 때 검사의 신 뢰도는 증가하는데 이는 Spearman-Brown 공식에 의한다. 검사의 길이와 신뢰도 계수의 관계를 제시하면 <표 3>와 같다.

### 3. 2. 2 문항의 변별도와 신뢰도

문항이 피험자를 능력에 따라 구분할 수 있

는 문항변별력이 있어야 검사의 신뢰도가 높 아진다. 안창규(1990)는 신뢰도는 문항 변별도 와 문항수에 직접적인 영향을 받으며 이들이 증가할수록 신뢰도도 증가한다고 밝히고 있다. 또한 신뢰도 증가에 영향을 주는 변인은 문항 변별도, 문항수, 문항점수 분산의 순이며 문항 수의 효과를 배제하였을 때 문항변별도가 다 른 변인에 비하여 신뢰도에 많은 영향을 주고 있음을 밝혔다.

### 3. 2. 3 문항의 난이도와 신뢰도

문항의 난이도가 적절하여야 신뢰도가 증가 한다. 검사가 너무 어렵거나 쉬우면 검사불안 과 부주의가 발생하여 진짜 능력을 발휘하지 못하므로 신뢰도가 저해된다. 또한 점수의 분 산은 문항난이도가 척도의 평균에 가까울수록 증가하기 때문에 난이도의 평균이 .5에 가까울 수록 신뢰도가 증가한다(안창규 1990). 이는 검사가 중간난이도를 가질 때 신뢰도가 증가 한다는 의미이다. 그러므로 문항을 선택함에 있어 문항 난이도가 중간수준이며 변별도가 높은 문항들을 선정할 경우 신뢰도는 증가하 고, 문항수가 많을수록 신뢰도는 높아진다.

### 3. 2. 4 검사문항의 내용과 신뢰도

측정도구의 측정내용이 보다 좁은 범위의 내용이어야 한다. 만약 인터넷 정보검색사 능 력시험을 치를 때 검사의 내용범위가 인터넷 네트워크로 제한된다면 인터넷 정보검색 전체

<표 3> 문항수와 신뢰도 계수

문항수	10	20	40	80	160	320	640	∞
신뢰도	.33	.50	.67	.80	.89	.94	.97	1.00

의 내용을 포함하는 검사보다 신뢰도가 높을 것이다. 이는 검사 내용의 범위가 좁을 때 문항간의 동질성을 유지하기가 용이하기 때문이다.

### 3. 3 타당도와 신뢰도의 관계

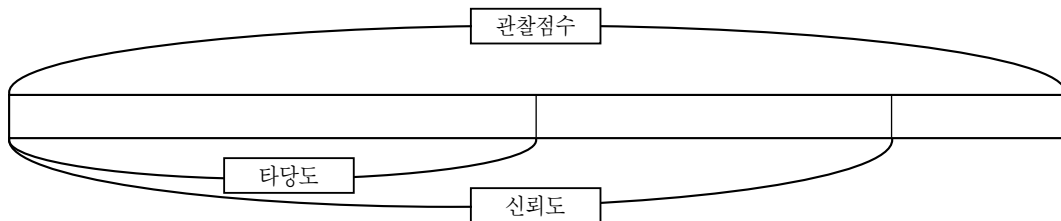
타당도는 검사문항이 측정하고자 하는 내용을 얼마나 충실히 측정하였느냐 하는 검사목적에 따른 검사문항의 적합성이고, 신뢰도는 측정하고자 하는 내용을 얼마나 정확하게 오차 없이 측정하였느냐 하는 검사문항의 일관성을 말한다. 측정점수를 구분한다면 진점수와 오차점수로 구분할 수 있고, 진점수는 타당한 진점수와 타당하지 않은 진점수로 구분되며 <그림 1>과 같다.

검사문항에 의한 관찰점수는 크게 진점수와 오차점수로 구분되어 신뢰도는 진점수에 해당되는 부분을 말한다. 진점수는 검사문항의 특성상 측정하고자하는 내용을 측정한 타당한 점수와 다른 특성을 측정한 타당하지 않은 점수로 구분된다. 타당도는 전체 관찰 점수 중 타당한 점수부분이 타당도가 된다. 그러므로 타당도가 높기 위하여 신뢰도는 높아야 한다. 그러나 신뢰도가 높다하여 타당도가 항상 높은 것은 아니다. 신뢰도와 타당도의 관계에서

신뢰도는 타당 도를 위한 필수조건이지 충분 조건은 아니다. 또한 타당도는 신뢰도의 제공 근 값보다 작거나 같다.

### 4. 문헌정보학 연구에서 타당도 및 신뢰도 적용시 고려해야 할 요소

이상에서 타당도 및 신뢰도에 대하여 살펴 보았다. 먼저, 내용타당도, 구인타당도, 준거타당도에 대해 살펴보았는데 이들 모두는 개념적으로 독립되어 있으므로 각각의 타당도는 측정도구의 다른 측면의 타당성을 검증한다. 예를 들면, 내용타당도는 모든 검사문항의 타당성을 검증하는 기본적인 절차로서 검사가 측정하고자 하는 내용을 측정하는지를 검증한다. 구인타당도는 측정하고자 하는 내용을 구인들로 규정하고 그 구인들을 제대로 측정하였는지를 검증하는 합리적인 방법으로 고려되어 사용경향이 증가하는 추세이다. 준거타당도와 관련하여 Kubiszyn과 Borich(1993)는 일반적으로 측정도구의 공인타당도가 예측타당도보다 높게 추정된다고 하였다. 이는 공인타당도는 동시에 추정되는데 비하여 예측타당도는 얼마간의 시간이 지난 후에 행위변수와



<그림 1> 타당도와 신뢰도의 관계

의 관계를 추정하기 때문이다. 공인타당도가 .8이상이거나 예측타당도가 .6이상이면 검사의 타당도가 있다고 할 수 있다.

타당도 계수는 피험자 집단의 응답결과에 의하여 추정되므로 피험자 집단이 보다 이질적일 때 공인 타당도와 예측타당도 계수가 높아지는 경향을 보인다. 반대로 피험자 집단이 동질적이면 타당도 계수는 낮아진다. 예를 들면, 대학수학능력시험의 예측타당도를 추정할 때 일류대학에서 예측타당도가 낮게 나온다. 해당대학에 입학하는 피험자들은 어느 일정수준 이상의 점수를 획득한 학생으로서 피험자 집단이 동질적이다. 그러므로 대학수학능력시험의 분산의 폭이 좁아지므로 대학학점과의 상관은 낮게 산출된다는 것이다.

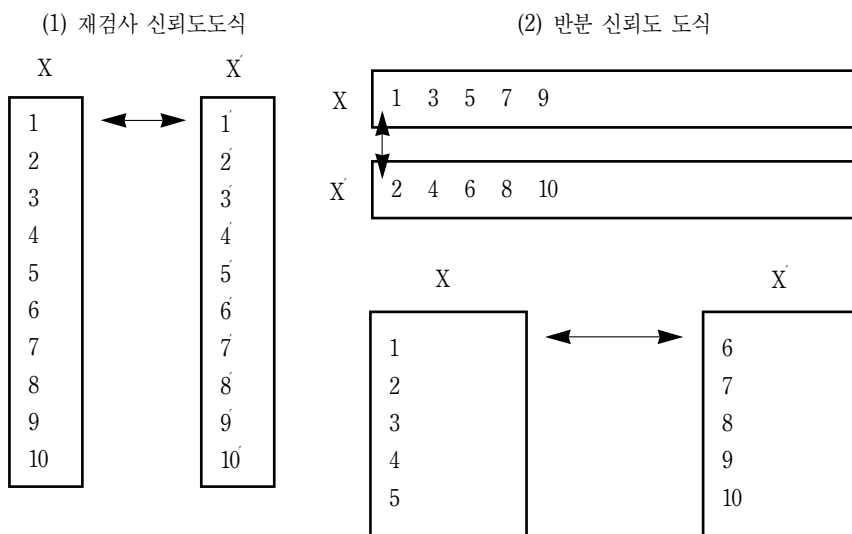
타당도 검증은 측정도구의 측정목적의 달성 여부를 증명하므로 모든 행동과학을 위한 자료분석의 기본절차이다. 타당도 검증 없이 얻어진 자료를 가지고 심리적 증상의 진단, 인성

측정, 학업성취도 비교, 교육효과 확인 등은 잘못된 결론을 유도할 수 있게 된다. 따라서 문헌정보학 제현상을 과학적으로 연구하기 위해 위에서 제시한 모든 방법에 의하여 타당도를 검증하는 것이 바람직할 것으로 보인다.

한편, 신뢰도와 관련해서 재검사 신뢰도는 동일한 검사를 피험자들에게 두 번 행하고 그 결과에 대하여 상관계수를 산출해서 구하는 신뢰도 추정치  $r_{xx'} = \rho_{xx'}$ 이다 (Lord & Novick, 1968). 만약 하나의 검사가 10개의 문항을 포함한다면 재검사와 반분신뢰도의 차이를 <표 4>와 같이 나타낼 수 있다.

<표 4>에 나타난바와 같이 재검사 신뢰도는 같은 문항들이 서로 대비(match)되어 상관계수가 구해진다. 그러므로 다소 이질적인 문항이 포함되더라도 T-동형(T-equivalent)이 유지된다. 그러나 반분신뢰도의 경우는 이질적인 문항이 포함되었을 때 T-동형을 유지하기 힘들다. 따라서 반분신뢰도를 사용하고자 하

<표 4> 신뢰도 차이에 대한 도식



는 경우에 어떤 검사의 문항들이 서로 다른 빈도수식어들을 포함하고 있다면 반분된 문항들 간의 T-동형을 유지하기 어렵게 될 것이며, 이러한 경우에 신뢰도는 낮아지게 된다. 따라서, 미래의 문헌정보학을 연구하는데 있어서 신뢰도 검증시 이러한 점을 고려해야 할 것으로 보인다.

## 5. 결 론

그 동안 문헌정보학을 연구하는 논문들이 양적으로 상당히 증가하였고 주제영역 또한 다양해졌다. 또한, 데이터 수집방법 및 데이터 분석방법들이 기존의 사회조사 및 역사적 연구방법에서 탈피하여 실험연구 및 기타 계량 분석방법들이 많이 적용되어 왔다. 이러한 시점에서 문헌정보학 제 현상을 과학적으로 연구하기 위해서는 타당도 및 신뢰도 검증이 필요하다. 그렇지만 아직까지 문헌정보학 관련 학술 논문들 중에는 타당도와 신뢰도에 관한 검증은 거의 이루어져 오지 못한 실정이다. 일반적으로 타당도와 신뢰도 검증 없이 연구결과를 도출한다는 것은 연구결과가 일반화되기 어렵고 과학적이고 객관적인 결론을 도출할 수 없다. 따라서 본 논문에서는 이러한 문제점을 해결하는데 기초가 되는 타당도와 신뢰도에 대해 살펴보았다.

타당도는 검사문항이 측정하고자 하는 내용을 얼마나 충실히 측정하였느냐 하는 검사목적에 따른 검사문항의 적합성이고, 신뢰도는 측정하고자 하는 내용을 얼마나 정확하게 오

차 없이 측정하였느냐 하는 검사문항의 일관성을 말한다.

타당도검증은 측정도구의 측정목적의 달성 여부를 증명하므로 모든 행동과학을 위한 자료분석의 기본절차이다. 타당도 검증 없이 얻어진 자료를 가지고 심리적 증상의 진단, 인성 측정, 학업성취도 비교, 교육효과 확인 등은 잘못된 결론을 유도할 수 있게 된다. 따라서 문헌정보학 제 현상을 과학적으로 연구하기 위해 위에서 제시한 내용타당도, 구인타당도, 준거타당도 등 모든 방법에 의하여 타당도를 검증하는 것이 바람직할 것으로 보인다.

신뢰도 검증은 타당도와 더불어 문헌정보학을 연구하는데 필요한 절차이다. 재검사 신뢰도는 같은 문항들이 서로 대비(match)되어 상관계수가 구해지므로 다소 이질적인 문항이 포함되더라도 유효한 검증이 가능하다. 그러나 반분신뢰도의 경우는 이질적인 문항이 포함되었을 때 반분된 문항들 간의 동형을 유지하기 어렵게 될 것이며, 이러한 경우에 신뢰도는 낮아지게 된다. 따라서, 이러한 점을 고려해서 신뢰도 검증방법을 선택하는 것이 바람직하다.

타당도와 신뢰도에 영향을 미치는 요인으로 는 여러 가지가 있지만 검사문항 제작과 관련해서는 빈도수식어, 문항진술방향, 문항의 난이도 및 변별도, 문항수 등이 있다. 따라서, 질문지 문항이나 기타 실험결과 테스트 문항을 작성할 경우 이상의 점들을 유의해서 작성할 경우 타당도와 신뢰도를 높일 수 있을 것으로 보인다.

## 참 고 문 헌

- 안창규. 1990. 검사의 신뢰도를 높이기 위한 문항선정기준에 관한 연구 『부산대학교 학생생활연구소』 연구보 제 25집
- 이두영, 김성희, 이명희. 1997. 『문헌정보학연구 방법론』 한국도서관협회.
- 정희태. 1987. 문항의 길이, 수식어, 진술방향이 문항반응분포 및 신뢰도와 타당도에 미치는 영향, 부산대학교 교육대학원, 『석사학위논문』 1987.
- Altermeyer, R. A. 1970. "Adverbs and interval: A study of Likert Scale". *Proceedings of the American Psychological Association*: 397-398.
- Anastasi, A. 1954. *Psychological Testing* [1954, 1968, 1976, 1982] 1988. New York: Macmillan.
- Cramer, W. 1951 "The positive or negative effect of the form of the question upon college students's answer". *Journal of Educational Research*,: 303-307.
- Cronbach, L. J. 1949. *Essentials of psychological testing*(1st ed). New York: Harper
- \_\_\_\_\_. "Coefficient alpha and the internal structure of test". *Psy-chometrika*, 17. :297-334
- \_\_\_\_\_. 1960. *Essentials of psychological testing*(2-5th ed). New York: Harper.
- Deaton, W. L. 1980. "Effects of item characteristics on Psycho-metric Properties of forced choice scales". *Education and Psychological Measurement*, :599-610.
- Edwards, A. L. 1957. *Techniques of attitude scale construction*, New York: Appleton-Century-Crofts.
- Kerlinger, F. N. 1984. *Foundation of Behavioral Research*. 2nd. ed., N.Y., Holt, Rinehart-Winston.
- Kubiszyn, T. & G. Borich. 1993. *Educational testing and measurement*, IL:Scott, Forcemen and Company.
- Lord, F. M. & M.R. Novick. 1968. *Statistical Theories of Mental Test Scores*, Menlo Park, California Addison-Welsley.
- Simpson, R. H. 1944. "The specific meanings of certain terms indicating degree of frequency". *The Quarterly Journal of Speech*.: 328-330.
- Vinacke W. et al., 1964. *Dimensions of Social Psychology*, Glenview, Ill. Scott, Foresman and Co., Preface
- Whipple, J. W. 1957. "A study of the extents answer in positive or negative phrasing affects answer in a true-false test". *Journal of Educational Research*: 59-63