

測定 및 測定道具의 妥當性과 信賴性

金 海 東*

<目 次>	
제 1 절 測定의 妥當度	3. 信賴度의 檢查 (1) 再検査法 (3) 複數樣式法 (3) 反復法
1. 意 義	4. 信賴度의 提高方法
2. 妥當度의 形態 (1) 論理的 妥當度 (2) 經驗的 妥當度 (3) 構造的 妥當度	제 3 절 妥當度와 信賴度의 相互 關係
3. 測定의 妥當化方法 (1) 論理的 妥當化法 (2) 專門家意見法 (3) 既知集團法 (4) 獨立基準法	제 4 절 妥當度와 信賴度에 영향을 미치는 要因 1. 檢查道具 및 그 內容 2. 環境的 要因 3. 個個人的 要因 4. 調查者의 解釋
제 2 절 信賴度	
1. 意 義	
2. 信賴度의 理論的 接近	

第一節 測定의 妥當度

1. 意 義

自然의 事物을 測定할 때에는 大部分의 境遇 測定의 對象을 直接的으로 測定하는 것이 普通이다. 그러나 社會科學의 對象을 測定할 때에는 反對로 大部分 그 自體의 性質上 間接的으로 測定하여야 하는 경우가 大部分이다. 이러한 事實 때문에 社會科學에서 妥當性的 問題가 重要視되는 것이다. 換言하면 社會現象을 測定할 때에는 조사자가 實際히 自己가 測定하고자 하는 바 이것을 測定하였다는 確信을 完全히 가질 수 없는 境遇가 많은 것이다. 即調查者가 測定하려고 하는 것을 測定하였느냐 하는 問題가 바로 妥當性的 問題인 것이다. 例를 들면 어떤 社會의 政治發展의 程度를 測定하려고 할 때 選舉演說集會에 모인 사람들의 數라든가 또는 投票率을 調查하여 그것을 測定하였다고 말할 수 있는가 하는 것이다.

一般的으로 조사자가 測定하고자 하는 바의 것을 測定하였느냐 하는 것이 妥當性的 問題

* 서울大學校 行政大學院 教授

이며 그러기 為하여는 測定道具가 妥當性이 있는 道具인지가 檢討되어야 하고 또한 그러한 測定道具의 適用方法이 妥當하여야 한다. 여기서는 主로 測定道具의 妥當性을 檢討하는 方法에 關하여 說明하기로 한다.

2. 妥當度의 形態

測定道具의 妥當度를 檢討하는 方法으로 普通 다음과 같은 세가지 方法이 있다. 첫째는 論理的方法(logical validation)이다. 이것을 外見上의 妥當性(face validity) 또는 內容上의 妥當性(content validity)이라고도 하며 主로 作成된 測定道具의 項目들이 論理的으로 調査者가 測定하고자 하는 內容을 包含하고 있느냐 하는 것을 檢討하는 것이다. 둘째는豫測의 妥當性(predictive validity) 또는 併行의 妥當性(concurrent validity)이라고 하는 것이며 主로 測定結果와 外的으로 存在하는 어떠한 變動과의 經驗的關係를 보는 것이다. 그래서 이 것을 經驗的妥當性(empirical validity) 또는 基準關聯妥當性(criterion-related validity) 또는 獨立基準法(Independent criteria)이라고도 한다. 셋째는 構造的妥當性(construct validity)으로서 測定道具를 그것이 關聯되고 있는 全體의in 理論的構造(theoretical framework)에 關聯시켜보는 것이다. 이제 이 세가지 方法을 좀더 仔細히 살펴 보기로 한다.

1) 論理的妥當性(logical validity)

前述한와 같이 이것을 內容上의 妥當性(content validity) 또는 外見上의 妥當性(face validity)라고도 부른다.

이것은 尺度가 測定하고자 하는 對象의 內容을 얼마나 包含하였는가를 論理的으로 따져보는 것이다. 換言하면 測定道具가 測定對象이 가지고 있는 重要한 屬性들 中의 一部를 代表性 있게 包含하고 있으면 그 尺度는 論理的妥當性 또는 外見上의 妥當性이 있다고 보는 것이다.

例를 들면 學生들에게서 가끔 期末試驗에 「끼상한 問題가 나왔다」고 不平하는 소리를 듣는다. 이것은 그가 學期에 그 科目에 對하여 功夫한 內容中에서一般的으로 期待하지 않았던 試驗問題가 出題되었다는 뜻이다. 이것은 다시 그 學科에 關하여 그들이 功夫한一般的인 內容이 아니라 아주 特殊한 問題라는 뜻이기도 하다. 換言하면 그 科目에 對한 試驗問題은 그 科目全般에 걸쳐서 어느程度 理解하고 있느냐하는 것을 檢討할 수 있는 것이 못되었다는 것이다. 勿論 學校에서는 試驗의 目的이 學習의 延長的性格도 있고 있지만 이러한面을 考慮한 出題가 아니라 그 科目的 理解程度를 알기 為한 것이라면 그 試驗의 結果를 가지고 그 理解程度를 推定할 수 있는 것이어야 되며 그러기 為하여는 試驗問題의 選擇에 있어서 代表性이라는 點이 考慮되어야 한다는 것이다. 換言하면 그 科目的 特定한 部分 또는 問題가 選定되어서는 學生들의 그 科目에 對한 理解程度가 測定되지 않는다는 것이다.

그런데 實際로는 어떠한 態度를 測定하고자 할때에 그 態度를 反映하는 項目은 無數하게 存在한다고 假定할 수 있다. 이것은 마치 行政學概論이라는 科目에 關한 試驗問題를 數 없

이 想定할 수 있는 것과 같다.

그리하여 이러한 測定對象이 되는 態度次元이나 屬性 또는 어떠한 問題를 反映하는 項目의 母集團을 論理的으로 想定할 수 있으며 尺度에는 이와 같은 項目的 母集團(universe of items)에서 論理的으로, 그리고 可能하면 體系的으로 몇 個의 項目을 抽出(sampling)하여 利用된 것이며 理想의이라고 말할 수 있다. 換言하면 抽出된 項目들이 無限한 項目的 母集團을 적절히 代表할 수 있다면 論理的妥當度가 높다고 할 수 있는 것이다. 다시 말해서 項目的 母集團(u)로부터 適切한 項目을 표본추출하여 만든 測定道具는 가능한 한 u 로부터 많은 項目를 無作爲의으로 抽出해야 妥當度가 높은 것이다. 그러나 불행하게도 일반적으로 項目的 母集團이 實제로 존재하는 것이 아니라 理論的으로만 존재하는 것이므로 u 로부터 項目을 無作爲로 抽出하는 것이 불가능하다. 따라서 가능한 한 관련된 項目을 많이 수집하게 되며 이 수집된 항목에서 표본추출하게 되는 것이다. 그러나 이렇게 수집된 항목은 그것이 아무리 많고 또 훌륭한 것이라 하더라도 항상 論理的妥當度가 문제 되는 것이다.

그러므로 項目的 母集團을 標本項目이 어느 정도 잘 代表하고 있느냐 하는 것은 調查者의 主觀的判斷에 依存할 수 밖에 없는 것이다. 다시 말해서 調查者에 따라 代表值에 대한 評價가 달라지기 쉽기 때문에 論理的妥當度는 대단히 主觀의이라 하지 않을 수 없는 것이다. 가령 人間의 疏外에 對한 研究를 한다고 할 때, 疏外의 概念은 여러가지로 생각해 볼 수 있다. 単に, 權力喪失狀態, 無意味, 無規範狀態, 孤立狀態, 冷談, 不和, 아노미等等을 들 수 있다. 이렇게 많은 概念들이 疏外란 概念을 전부 다 망라한 것인지, 또는 이 概念들 중 어떤 概念이 疏外를 가장 잘 代表하고 있는지를 일률적으로 규정할 수는 없는 것이며 이것은 調查者的 主觀的判斷에 依存할 수 밖에 없는 것이다. 이와 같이 論理的妥當度는 조사자의 판斷의 質에 달려있으므로 이 主觀的判斷이 合理의일 필요가 있는데 이를 위해서는 分析하여 현상에 대한 意味의 適用可能性을 추적하는 것이 필요하다. 또한 測定對象과 關聯된 知識 및 理論을 判斷의 基準으로 사용해야 하고 필요하다면 關係專門家의 意見이나 既存測定道具를 참고로 해야 할 것이다. 그리고 또 項目的 母集團을 가능한 한 분명하게 정의하도록 하는 한편 判斷의 內容과 方向을 구체화하도록 해야 할 것이다.

이제까지 說明한 測定道具의 內容의 代表性를 높이기 위해서는 다음과 같은 節次를 따를 것을 G.W. Bohrnstedt는 권하고 있다.

첫째, 어떠한 概念의 모든 意味를 전부다 포괄하기 위하여 중요하고 또 명백한 의미들을 열거한다.

둘째, 어떤 意味가 單一次元을 대표하지 못하면 이를 다시 세분하도록 한다. 예를 들어 權力喪失狀態는 政治的經濟的 및 가정에서의 권력상실상태로 분해 볼 수 있다.

셋째, 각각의 意味에 대하여 적어도 7~10개 항목을 사용하도록 한다. 이렇게 해야만 각각의 意味를 보다 잘 파악할 수 있을 뿐만 아니라 信賴度 또한 높아진다.

넷째, 선택된項目들이 때로는 예상했던대로 기능을 하지 못할 경우가 있으므로, 이런 항목은 제거해야 한다. 가령 몇개의次元을 측정하는項目이라든가 원하는 차원을 측정하지 못하는項目과 같은 것이다.

이상에서 살펴본 바와 같이論理的妥當度는調查者的主觀的인解釋과判斷에지나치게의존함으로써判斷에오차또는착오가개입할여지가많으며또한統計的檢證이이루어지지않는결점이있다.또측정하려고하는속성과이속성을반영하는項目間의相應關係(correspondence)의정도를파악할방법이없다.따라서이論理的타당도는측정의타당도를수립하는방법중가장신뢰도가떨어지는방법이라하지않을수없다(Black & Champion, 1976, p. 233)

반면 이것은統計的節次를 사용하지 않고 직접 적용될 수 있으며 또 질문 내용에 기초를 두어推定될 수 있고,推定될 수 있는 장점이 있다. 그리고 또適用이 용이하고 시간이 많이 소요되지 않는 이점이 있다. 즉 조사자가 많은 수의 항목을 대체하는 調查表를作成하려고 할 경우보다 정교한妥當化節次를 이러한項目들에 사용하려면限定된調查의資源시간자료등의상당한부분을지출해야한다. 그러나論理的妥當度에관심을한정시킨다면이러한資源을조사요원의훈련등에더많이충당할수있을것이다.

(2) 經驗的妥當度

經驗的妥當度(empirical validity)는基準關聯妥當度(criterion-related validity)라고도하는데豫測的妥當度(predictive validity)와一致的妥當度(concurrent validity)로구별해서사용하기도한다.

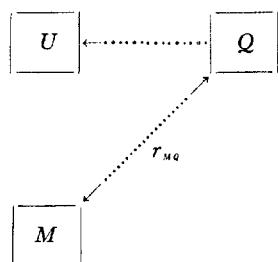
經驗的妥當度는統計적으로타당도를評價하는것으로測定道具와測定結果間의關係에관심을드는타당도이다.즉어떤측정도구가타당도가높다고할려면측정도구에의해생산된結果와다른어떠한基準또는變數間에經驗的인關係가存在해야하는것이다.다시말해서연구하려는屬性을측정해줄것으로알려진外의變數(基準)와測定道具의測定結果(尺度의點數)間의관계를비교함으로써타당도를파악하는방법이다.여기서關係가經驗의으로存在한다는證據는測定水準에맞는相關關係를측정함으로써구할수있다.이러한經驗的妥當度를評價하는方法에는여러가지가있으나豫測的妥當度와一致的妥當度가일반적으로가장많이사용된다.

豫測的妥當度는어떤검사가무슨행위가일어날것이라고예측한것과실제대상자또는집단이나타낸행위간의관계를측정하는것이다.예를들어몇몇會社의人事擔當職원들에대한男性優越主義를측정하는검사를한다고하자.그결과몇몇직원은남성우월적인태도를매우높게가지고있고,또몇몇직원은그러한태도를거의보이지않는것으로나타낸다고할때이경우대상자의남성우월주의에대한점수와실제그가보여준행위간의관계를비교함으로써측정의타당도가있느냐없느냐를결정하는것이다.즉남

성우월적인 태도를 높게 가진 직원이採用에 있어서 여성에 대한 性差別을 심하게 한다면, 또 남성우월적인 태도를 보이지 않는 직원이 성차별을 하지 않는다면 이 남성우월주의에 대한 측정은豫測的妥當度가 있다고 할 수 있는 것이다.理解를 돋기 위해 예를 하나 더 들어 보자, 먼저 대학신입생과 같은 어떤對象에知能検査를 실시하여 점수를 구한다음, 이들 신입생이 일년 동안에 받은評點의平均과 비교하여 이들간에 상관관계가 높으면 그知能検査는豫測的妥當度가 높다고 할 수 있는 것이다. 말하자면 두測定의點數를 계산하여相關係數 즉妥當度係數(validity coefficient)를 구하는 것이다. 이때 이妥當度係數가 높으면 예측적타당도가 높은 것을 의미하는 것이다.

測定道具의豫測的妥當度를評價하는過程은 어떤測定道具(M)에 의하여 어떤屬性(u)을測定한 다음 어떠한 명백한基準(Q)을 사용하여 이를비교하는 것이다. 즉 M 에 의해 획득된測定值와 Q 에 의해 획득된測定值의相互關係를계산한妥當度關係數(r_{Mq})의크기가測定道具의豫測的妥當度를測定하는 것이다.〈圖-1〉

圖 1.豫測的妥當度의評價過程



그런데豫測的妥當度를 사용할 때는 다음과 같은 사항이 검토되어야 한다.

첫째, 關聯基準의妥當度 문제이다. 즉基準을 확보하는 것도 어려울 뿐만 아니라 사용되는 기준의 타당도가 문제되는 것이다. 우리는 앞에서 그基準은 타당도가 있는 것으로 가정하였다. 이렇게 기준의 타당도가 문제되면 다시 새로운 다른 기준을 사용하여 타당도를 평가해보아야 한다. 이러한 과정은 타당도가 있다고 검증될 때까지 무한히 계속되어야 하는 것이다. 그런데 이러한基準의妥當度를 확보하는 방법으로 보통 다음과 같은 두 가지 방법이 사용되고 있다. 하나는 어떤 기준이 측정도구를 평가하는데 타당도가 있다고 조사자들 간에意見一致를 보는 방법이다. 여기서一致는論理的妥當度와關聯된다. 다른 하나의 방법은 측정도구와關聯基準과의關係를 후술할既知集團法(Known groups)에 따라 측정도구에 의하여 정확히 분류될 수 있는對象(또는事例)의比率로表現하는 방법이다.(Helmstadt er, 1970).

둘째, 예측적 타당도는 측정도구가 어떤屬性을 완전히 측정했다는 것을 증명할 수 없다는 것이다.

가령 政治的 保守主義를 측정하는 測定道具의 타당도를 평가한다고 하자. 이때 下位階層이 中間階層보다 保守的이라는 주장에 대한 이론적으로 충분한 근거를 가지고 있다면, 이 두 집단을 비교함으로써 예측적 타당도를 검토할 수 있을 것이다. 그런데, 이 경우에 社會的 階層은豫測的 妥當度의 間接的인 基準으로 作用하게 된다. 만약 이에 下位階層의 사람들이 中間階層의 사람들과 똑같이 保守的이라고 한다면 측정도구는 예측적 타당도가 결여되어 있다고 할 수 있다. 반대로 社會的 階層과 保守主義間에 상대적으로 높은 相關關係가 있으면 이것은 측정도구의 타당도를 높혀줄 것이다. 그러나 높은 相關關係는 측정도구의 예측적 타당도의 必要條件이긴 하나 充分條件은 아님을 유의해야 한다. 왜냐하면 社會的 階層과 같은 間接的 基準은 政治的 保守主義 이외의 다른 屬性과 관련될 수도 있기 때문이다 따라서 測定道具는 정치적 보수주의 대신에 다른 속성을 측정할지도 모른다. 이렇게 되면 간접적 기준은 측정도구의 타당도를 높히기 보다 오히려 타당도를 떨어뜨리는 작용을 하게 되는 것이다.

셋째 이때한 기준은 기술적으로 사용하기가 곤란하고 또 지나치게 비용이 많이 소요되기도 한다. 그리고 또 어떤 경우에는 기준으로 사용하기 이전에 그 속성을 측정해야 할 때도 있다.

넷째,豫測的 妥當度의 基準의 保守性에 관한 문제이다. 현재 사용되고 있는 基準이 쓸모없는 것임에도 불구하고 새로운 측정을 평가하는데 사용된다면 知識의 發展을 기대하기가 어렵기 되는 것이다. (Phillips, 1971, p. 199)

이제까지 설명한豫測的 妥當度의 問題點에도 불구하고 이것은 대부분의 경우에 타당도를 결정하는데 있어 가장 신뢰성 있는 방법이며 (Black & Champion, 1976, p. 233) 또한 適用이 用易하다.

다음 ---致的 妥當度는豫測的 妥當度와 時間的 次元에서 差異가 있는 것으로 말하자면豫測的 行爲의 點數가 실제 실현된 行爲와 동시에 획득된다는 점이 다른 것이다. 그러므로 일정기간 동안 應答者的 行爲를 검토할 필요가 없고, 따라서 예측적 타당도에 비해 보다 신속하게 適用할 수 있는 이점이 있다.

(3) 構造的 妥當度

構造的 妥當度는 精神測定的 觀念(psychometric notions)과 理論的 觀念을 결합한 것이기 때문에 現代測定理論 및 實際에 있어서 가장 괄목할만한 진전을 보인 분야 중의 하나이다. 이것은 外的 基準을 活用할 수 없는 屬性를 測定하는데 주로 사용되는 것이다. (Cronbach & Meehl, 1955) 다시 말해서 傳統的인 妥當化方法이 不適切할 경우에 사용된다. 즉 個性 檢查나 몇몇 能力檢査에서는 사용할 수 있는 적절한 기준을 찾기가 어렵다.

構造的 妥當度는 測定道具의 妥當度를 評價하는 手段으로 論理的 및 統計的 檢證을 使用한다. 이 점에서 論理的 妥當度가 論理的 檢證을 經驗的 妥當度가 統計的 檢證을 使用하는

것과 구별된다. 말하자면 이 妥當度는 단순히 測定自體의 妥當度를 문제삼기보다 그 測定의 기초를 이루고 있는 理論 및 理論的 構造와 假說의 關係의 檢證을 포함하는 科學的 經驗的 質問의 確保에 더 많은 관심을 집중하는 방식이다.

이를 보다. 구체적으로 설명하면, 構造的 妥當度는 測定道具를 全盤的인 理論的 構造 (overall theoretical framework)와 연결시켜 이것이 概念 및 理論的 前提와 어떠한 關係를 이루고 있는가를 규명함으로써 타당도를 파악하는 방법이다.

여기서 構念의 意義은 어떤 檢查의 施行結果 얻어진 점수는 무엇을 意味하는지 또 대상자가 그러한 점수를 얻게 되는原因是 무엇인지를 理解하도록 해주는 데 있다. 그리고 測定하려고 하는 屬性에 대한 理論的 觀念은 研究者로 하여금 이 屬性과 다른 特定變數의 關係를 규명하도록 해준다. 그런데 測定道具가 構造的 妥當度가 있다고 하려면 이러한 關係를 調査研究者가 실제로 보여주어야 한다(統計的 또는 經驗的 檢證)

예를 들다. 다음과 같은 두 가지 測定, 즉 教育的 傳統主義에 관한 測定과 ‘훌륭한’ 선생의 특성을 어떻게 인식하는가에 대한 측정에 있어, 이를 간에 順相關이 있다고 하자. 다시 말해서 傳統主義에 관한 測定에서 높은 점수를 보인 사람은 ‘훌륭한’ 선생을 근면하고, 도덕적이고, 또 양심적이며, 철저하고 능률적인 사람으로 이해하는 반면 傳統主義 測定에서 점수가 낮은 사람들은 ‘훌륭한’ 선생을 다른 형태로 이해하고 있다고 하자. 여기서 조사자는 왜 이러한 關係가 존재하는 것이며, 이 關係가 意味하는 것은 무엇인가를 규명하고 싶어한다. 그런데 이러한 關係의 成立理由나 意味를 이해하기 위해서는 이 關係를 形成하는 變數 또는 構造 즉 「훌륭한 선생에 대한 認識」과 「傳統主義」의 意味를 알아야 하는 것이다. 構造的 妥當度란 말하자면 이러한 意味를 어떻게 연구하는가 하는 문제와 연결되는 것이다.

이러한 構造的 妥當度의 論理的 過程을 Cronbach와 Meehl (1955)은 다음과 같이 설명하고 있다. 첫째, 어떤 測定道具가 어떤 屬性 예를 들어 속성을 測定한다는 命題를 設定한다. 둘째, 이 命題를 屬性 “A”에 關한 理論에 포함시킨다. 셋째이 理論을 통하여 測定道具와 關聯되어야 할 다른 屬性와 관련이 있어서는 안될 또 다른 屬性를豫測한다. 끝으로 예측된 關係를 經驗的으로 確認해 주거나 거절하는 資料를 수집한다. 다시 말해서 構造的 妥當化는 試驗成績를 계산해주는 構造의 提示, 이러한 構造를 포함하는 理論으로부터 假說의 導出假說의 經驗的 檢證의 過程을 通해 구해지는 것이다.

여기서 그 대된 關係가 발견되면 그 測定道具는 妥當度가 있는 것으로 추정된다. 그런데豫測된 結果가 발견되지 않으면 이것은 다음 세 가지 경우로 생각해 볼 수 있다. 즉 ① 測定道具가 屬性를 測定하지 않았거나 ②豫測된 結果를 도출한 理論的 構造가 不正確하거나, 아니면 ③ 調査設計가豫測된 結果를 적절히 檢查하는데 실패 한 것이다. 調査者는 이러한 問題點이 발생했을 경우에는 妥當한 過程을 形成하는 세 가지 段階의 각각을 주의깊게 再構成해야 한다.

Campbell과 Fiske (1959)는構造的妥當化의 또 다른 방법으로相關行列(correlation matrices)을 들고 있다. 즉妥當設의收斂一差別概念(convergent-discriminant conception of validity). 多重屬性—多重方法行列技法(multitrait-multimethod matrix technique)이 그것이다. 이러한方法은測定方法이 달라도 동일한屬性을測定하는 것은同一한結果를산출하는반면相異한屬性은測定道具에 관계없이상이한測定結果를산출해야한다는論理에근거를둔것이다. 다시말해서이것의操作的인意味는상이한척도에의해측정된어떤하나의屬性에대한점수측정결과간의相關係數는상이한척도에의해측정된서로다른속성에대한점수간의相關係數보다높아야한다는것을말한다. 그러므로測定道具의構成的妥當度는다음두가지原理를사용하여그證據를확보할수있다. 즉收斂原理(convergent principle)과差別原理(discriminant principle)가그것이다. 여기서收斂原理란동일한속성에대한두測定은비록서로다른방법을사용한다고하더라도각각은높은상관관계를가져야한다는것을의미하며, 差別原理란유사한측정도구가사용하더라도상이한속성을측정하는것이면두측정간에서로높은상관관계가있어서는안된다는것을의미한다. 따라서妥當化過程은적어도두가지이상의屬性을代表하는측정도구의相互關係를계산할필요가있으며, 이를속성은각각두가지이상의測定道具로측정되어야하는것이다. 그런데이방법은대단히이상적이긴하나상대적으로많은표본에대해두개이상의變數를두번이상측정한다는것은대단히어려운일이다.

構造的妥當化的또다른방법으로調査에서사용되는방법을예로들어보자. 어떤의미에서는妥當化는그것이어떤형태이던간에構造的妥當化라고할수있다. 즉假說檢定이나關係의經驗的研究등에는構造的妥當度가내포되어있는것이다. 調査에서構造의妥當化를사용한예로는여러가지가있으나Rokeach의獨斷主義에관한調查로살펴보기로한다.

그는理論的論理에根據를두고獨斷主義에관한尺度를構成하였다. 이척도는폐쇄적성격(closed mindedness)과이데올로기와관련된思考의樣式을파악해줄것으로기대되는몇개의項目으로구성되어있다. 그는어떤개인이어떠한이데올로기를가지느냐는것은그의성격, 사고의과정, 행위와관계가있다고주장하면서독단주의가강한사람은자기의의견을고집하는사람이라는豫測을하였던것이다. 그리고또Rokeach는그의一聯의연구의목적을理論의檢證및測定道具의構造的妥當度에두었는데, 종교가서로다른집단을검정하면서카톨릭을믿는학생들이기독교를믿는학생들보다더독단적인것을발견하였다. 이것은그의豫測를뒷받침하는것이었다. 그리고또공산주의자가자유주의자보다더높은점수를얻은것도발견하였던것이다. 이러한研究의結果는명백한것은아니지만, 獨斷主義에관한尺度가構造的妥當度가있다는증거를제공해주는것이다.

構造的妥當化的또다른방법으로要因分析(factor analysis)를들수있는데이것이아마

가장 강력한 방법일 것이다(Kerlinger, 1973, p. 468) 간단히 말해서 要因分析은 測定의 기본적 구조를 결정하는 통계적 기법으로, 多數의 測定에 있어서 동일현상의 측정을 하나로 묶어만든 보다 적은 수의 측정 즉 要因으로 축소하여, 이를 測定의 集合들간의 관계를 측정하는 방법이다. 말하자면 요인분석은 어떠한 측정이 동일한 것을 측정하는가, 또 어떤 측정이 측정하려고 하는 것을 어느정도 측정해 주는가를 효과적으로 제시해 주는 것이다.

이상에서 살펴본 바와 같이 構造的 妥當度는 潜在的인 態度의 측정에 유용하긴 하나 직접 관찰할 수 있는 속성에 적용하는 것이 일반적이므로, 조사하려는 현상과 비슷한 다른 현상을 측정할 어렵지 있음을 유의해야 한다.

3. 測定道具의 妥當化 方法

測定道具가 有用하기 위해서는 妥當性을 인정할 수 있는 證據를 확보해야 하는데, 社會의 現象이나 人間의 行爲에 있어서 이러한 證據를 확보하기가 실제로는 대단히 어려운 문제이다. 그런데 이러한 측정도구를 妥當化(validation)하는 방법으로 完全하지는 않으나 다음과 같은 네 가지 방법이 사용되고 있다.(Goode & Hatt, 1952, pp. 237-239) 그런데 이러한 妥當化方法는 妥當度의 形態에서 이미 조금씩 설명한 바 있다. 또 어떤 의미에서는 妥當度 檢查方法의 形態가 바로 타당화방법이라고도 할 수 있는 점도 유의해야 할 것이다. 그럼에도 불구하고 이름 따로 설명하는 것은 이를 보다 체계적으로 설명할 필요가 있고 또 이를 妥當度를 높히는 方法으로도 理解할 수 있기 때문이다.

(1) 論理的妥當化法(logical validation)

이것은 가장 널리 사용되는 방법으로 外觀的 妥當化(face validation)라고도 한다. 이것은 測定道具의 項目을 理論的 또는 常識的으로 分析하여 타당도를 확보하는 방법으로, 測定하고자 하는 事項을 정확하게 규정하고 항목을 조심성 있게 선택함으로써 자동적으로 이루어진다 예를 들어 保守主義를 測定하고자하는 測定道具의 項目에는 財產·家族, 또는 政治制度等에 대한 態度를 파악하는 質問이 포함되어 있을 것이며, 이러한 질문에 대한 態度表示를 상식적으로 판단하게 되는 것이다. 즉 現狀維持를 지지하는 태도는 保守的이라고 하는 극히 常識的인 판단을 하게 되는 것이다.

그러나 이처럼 論理的 妥當화가 그렇게 쉬운 일이 아니다. 保守主義에 관한 測定項目에는 共產主義에 대한 質問이 포함될 수 있으며, 이때 상식적으로는 보수주의적인 사람들은 공산주의를 대단히 싫어하고, 自由主義者들은 보다 관대한 태도를 보일 것이라고 생각할 수 있다. 그러나 실제로는 保守主義者들이 個人的 権利에 대한 法的保障이라는 19세기적 사상을 견지하는 입장에서 공산주의 자들을 억제하는 것을 좋지 않게 생각할 수도 있으며, 또 자자유주의자들도 그들이 생각하는 急進的인 社會改革運動을 배반했다고 하여 共產主義者를 보호할 필요가 없다고 주장할지도 모르는 것이다.

따라서 測定道具의 妥當化를 全的으로 論理的 또는 常識的인 방식에만 의존한다는 것은 위험한 일이라 아니할 수 없다.

(2) 專門家意見法(jury opinion)

이것은 어떤 意味에서 論理的 妥當化의 延長이라고 할 수 있는 것으로, 測定道具가 測定하고자 하는 문제에 관한 전문가로 생각되는 사람들의 論理的인 判斷에 의거하여 타당도를 확보하는 방법이다. 예를 들어 住宅의 適合程度를 측정하기 위한 測定道具를 만들 때 기술자, 건축가, 복덕방업자, 주부들의 판단에 의하여 타당도를 결정하는 것이다. 또 척도의 구성에 있어서 이들 關係專門家의 意見을 참고함으로써 타당도를 높일 수도 있는 것이다. 그런데 이 방법은 전문가의 論理的 判斷에 근거한다는 것 외에는 論理的 妥當化와 차이가 없다고 할 수 있다.

(3) 既知集團法(known groups)

이것은 專門家意見法을 약간改善한 것으로, 이미 그 意度와 特徵을 알고 있는 集團에 測定道具를 적용하여 이들의 태도와 특징을 충분히 식별해내면 그 測定道具는 타당도가 있는 것으로 評價하는 방법이다. 다시 말해서 상반된 태도와 특징을 가진 것으로 이미 알려진 두 개의 집단을 설정하여 이들에 측정도구를 적용한 결과 그 結果의 差가 상반된 태도를 나타내는지를 알아보는 방법이다. 예를 들어 土氣를 측정하는 測定道具를 구성할 경우 사기가 높은 집단과 사기가 낮은 집단에 동일한 측정도구를 적용하여 두 결과의 차이를 보는 것이다. 이때 차이가 분명하게 나타나면 타당도가 있는 것이요, 차이가 나타나지 않으면 그 측정도구의 타당도는 없다고 할 수 있는 것이다. 즉, 陸軍將校補職管理處에서는 作成된 測定道具의 타당도를 검사하기 위해 進級에 1회 누락된 사람, 2회 누락된 사람, 3회, 4회 누락된 사람들과 한번도 누락되지 않은 장교들을 대상으로 이 척도를 적용하였다. 그 結果 진급에 누락된 회수가 많은 장교일수록 土氣점수가 낮았다. 따라서 이 측정도구는 타당성이 있다고 인정할 수 있는 것이다.

이 방법은 비교적 널리 사용되나 항상 다음과 같은 문제점을 내포하고 있다. 즉, 두 集團間에 土氣 이외의 다른 태도나 특징이 차이가 있어 이것들이 結果의 차이에 영향을 미친 것인지 알 수 없다는 것이다. 예를 들어 사기에 관한 차이외에도 연령, 학력, 결혼상태, 가정적환경, 軍에 대한 태도, 각종 취미 등의 차이가 있어서 結果의 차이가 나타난지도 알 수 없다는 것이다. 이러한 차이가 사기와 아무리 상관관계가 높다고 하더라도 그것은 근본적으로 사기를 측정하는 것과는 다른 것이다. 그러므로 진급에 누락된 회수를 사기점수상으로 식별할 수 있다고 하여 土氣尺度가 타당성이 있다고 인정할 수는 없는 것이다. 따라서 既知集團法은 상당한 주의를 요한다 특히 이러한 妥當化方法은 단지 그럴 듯하다는 것 이지 결코 타당도를 확실하게 증명해주는 것은 아니라는 점을 망각해서는 안된다. 그럼에도 불구하고 이 방법이 많이 사용되고 있는 것은 이러한 高度의 近似度(hight degree of pla

usibility)이 외에는 별다른 방도가 없기 때문이라고 하는 것이 옳을 것이다.

(4) 獨立基準法(independent Criteria)

이것은 이론적으로 볼 때 가장 이상적인 방법이나 실제 적용하는데 있어서는 대단히 어려운 방법이다. 이 방법은 어떠한 獨立的인 기준에 입각하여 그 측정도구의 타당도를 검사하는 방법이다. 그런데, 측정하고자 하는 대상에 대한 독립적인 한개의 기준이 존재하고 있다면 사실상 새로운 측정도구를 구성할 필요성이 하나도 없는 것이다. 따라서 보통 그 측정도구의 기초가 되는 몇 가지 기준을 사용하여 그 측정도구의 타당도를 검사하는 것이다. 그리하여 그 측정도구는 그것만으로써 보다 더 측정하고자 하는 바를 용이하게 측정하도록 해주는 것이다.

예를 들어 社會的基盤(social standing)을 측정하는 척도를 작성하고자 한다면 그 척도의 타당도는 이것을 주택을 買를 놓았을 때의 가격, 교육수준, 收入, 가정적배경, 재산, 직업 등의 여러가지 요인과 대조하여 판단하는 것이다. 이것은 이러한 여러가지 요인의 總和가 사회적기반을 올바르게 측정해 줄 것이라는 가정에 입각한 것으로, 이러한 요인들과 측정도구가 어느정도가 부합되느냐 하는 相關關係가 그 척도의 타당도를 나타내는 것이라고 하는 것이다.

그러나 이러한 利用可能한 獨立的인 基準이 존재하지 않을지도 모르며, 또 그것이 존재하여 측정하고자 하는 바를 잘 반영시켜준다면 구태여 측정도구를 작성할 필요가 없는 것이다. 그리고 또 이러한 獨立的인 基準自體가 척도가 측정하고자 하는 바의 것에 대한 적절한 지표가 되지 못할지도 모른다. 이러한 문제만 해결된다면 이 獨立基準法은 가장 효과적인 방법이 될 수 있다.

일반적으로 尺度道具의 妥當化를 위해서는 이상 네 가지 방법을 가능한 한 모두 적용하는 것이 좋으며 特히 論理的 妥當化法은 반드시 채용하도록 하는 것이 좋다. 그러나 이것만으로는 불충분하므로 기타 다른 방법중 하나 이상을 並用하여 적용하는 것이 좋다.

第2節 信賴度

1. 意義

信賴度(reliability)는 測定道具가 측정하려고 하는 현상을 일관성있게 측정하는 能力を 말한다(Sellitz, et al. 1959). 보다 일반적으로 설명하면 어떤 測定道具를同一한 現象에 反復適用하여 同一한 結果를 얻게 되는 정도를 그 측정도구의 신뢰도라 한다.

이러한 信賴度의 價値는 크게 다음 두가지로 생각해 볼 수 있다.

첫째, 信賴度는 妥當度를 확보하기 위한 필수적인 前提條件일 뿐만 아니라 妥當度보다 상대적으로 확보하기가 용이한 것이다.

둘째, 信賴度는 수집된 資料에서 나타난 差異가 測定道具 自體의 不完全에서 結果되는

것이 아니을 확신케 해준다.

말하자면 信賴度는 測定에 있어서 가장 중요한 것은 아니지만 그래도 아직까지는 매우 중요하다고 하지 않을 수 없다. 다시 말해서 신뢰도가 높다고 해서 훌륭한 科學的 結果를 보장하는 것은 아니지만 신뢰도가 없는 훌륭한 과학적 결과는 존재할 수가 없는 것이다. 간단히 이 얘기해서 신뢰도는 研究調查結果와 그 解釋에 있어 必要條件이나 充分條件은 아니라는 것이다.

이러한 신뢰도와 유사하게 사용되는 말로 信憑性(dependability), 安定性(stability), 一貫性(consistency), 豫測性(predictability), 正確性(accuracy) 等이 있다. 예를 들어 신뢰성이 있는 사람이라고 하면 그의 행동이 일관되고 신빙성이 있으며 예측이 가능한 사람을 말하는 것이다. 다시 말해서 앞으로 그가 취할 행동은 그의 지난날의 행동과 일관되며 또, 현재 그의 행동으로 미루어 보아 앞으로 그가 어떠한 행동을 취할지 예측할 수 있을 때 그를 신뢰성이 있는 사람이라 할 수 있는 것이다.

이러한 信賴度를 定義하는 方法은 測定의 어떤 측면을 강조하느냐에 따라 다음 세 가지 接近方法으로 나누어 생각해 볼 수 있다,

첫째, 신뢰도를 동일한 대상을 동일한 측정도구를 사용하여 측정할 경우 동일한 결과를 얻을 수 있는 정도로 정의하는 방법이다. 이러한 신뢰도의 定義에는 安定性, 信憑性, 豫測性의 의미가 크게 내포되어 있다. 이 概念定義方法은 기초적인 定義로서 가장 많이 사용되는 방법이다.

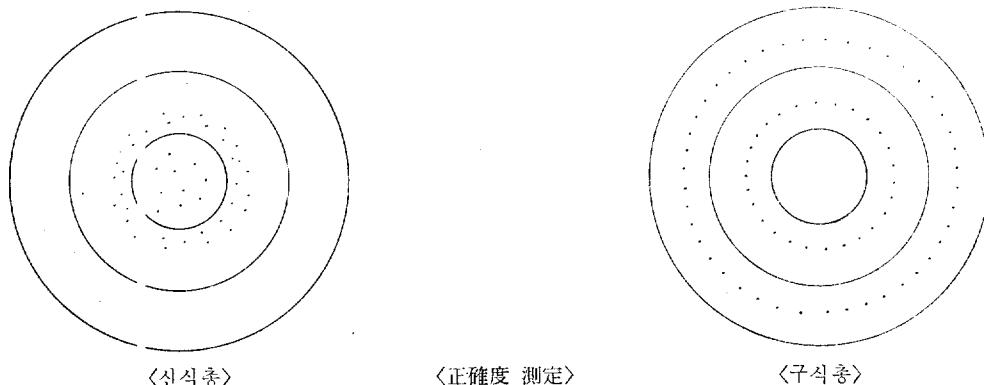
둘째, 正確性 및 精密性 즉 측정하고 하는 屬性을 얼마나 진실에 가깝도록 측정하느냐 하는 것으로 신뢰도를 정의하는 방법이다. 이러한 正確性 및 精密性을 강조하는 정의는 첫번째의 安定性을 강조하는 定義를 내포하고 있다.

셋째, 測定道具에 있어 測定誤差(error of measurement)가 어느 정도인가를 파악함으로서 신뢰도를 정의하는 방법이다. 이 방법은 신뢰도의 실제적인 문제점과 이론적인 문제점을 보다 잘 해결해 주고 또 개념정의를 분명하게 해줄 뿐만 아니라 다른 개념정의나 접근방법을 팔적으로 설명하게 해준다.

즉, 誤差에는 體系的 誤差(systematic error)와 無作爲誤差(random error)가 있는데 신뢰도에 관한 오차라 함은 測定誤差 즉 無作爲誤差를 말한다. 여기서 體系的 誤差란 測定過程 또는 測定對象에 대하여 어떠한 영향이 체계적으로 미침으로서 발생하는 오차를 말하며 이것은 測定結果가 모두 공정적이 아니면 모두 부정적이거나, 모두 높거나 모두 낮은 경향을 보이는 것이 특색이다. 다시 말해서 체계적 오차는 誤差가 일정하거나 또는 한쪽으로 치우쳐(biased) 있다. 반면 無作爲誤差는 測定者의 피로, 記憶 또는 감정의 變動等과 같이 測定對象, 測定過程, 測定手段, 測定者 등에 우연히 그리고 일시적으로 영향을 미침으로서 발생하는 오차를 말하며 이것은 人爲的인 것도 아니고 또 체계적인 것도 아니므로 誤差의

의 값이 다양하게 분산되어 있으며, 相互相殺되는 傾向(self-compensating)도 있다. 測定道具에 있어서 이러한測定誤差가 많으면 많을수록 그만큼 測定道具의 信賴度는 떨어지는 것이다. 다시 말해서 신뢰도란 측정도구에 측정오차가 상대적으로 얼마나 있는가 하는 것으로 정의할 수도 있다.

그럼 이제 신뢰도에 대한 두번째 정의 즉 測定道具의 正確性 또는 精密性에 대해서 보다 구체적으로 살펴보기로 하자. 먼저 정확도란 체계적 오차와 무작위오차를 합한 것의 大小를 말한다. 가령 어떤 스포츠맨이 두 銃의 정확도를 비교하고자 하는데 하나는 성능은 좋으나 1세기전에 만들어진 것이고, 다른 것은 전문적인 총제조 업자가 만든 것이라 하자 여기서 우리는 舊式銃의 彈痕이 新式銃에 비해 보다 넓은 범위에 걸쳐 흩어져 있음을 볼 수 있다. 그런대 표적에 있어서 중심원은 3. 맨바깥 원은 1, 그사이 원은 2, 그리고 표적에 맞지 않은 것은 0이란 숫자를 부여한다고 하자, 그리하여 두 銃의 彈痕을 가지고 散布度, 이를테면 표적편차를 계산한다고 하자. 그러면 신식총의 산포도가 구식총에 비해 적게 나

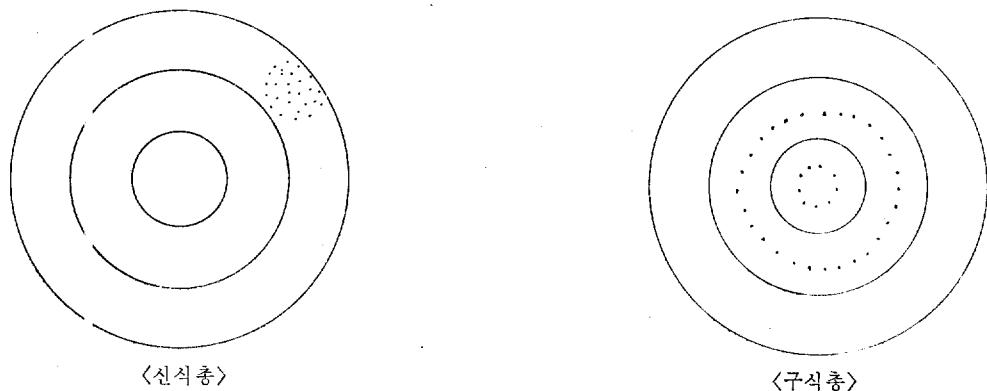


타날 것이 분명한데, 이것은 오차가 상대적으로 적어 정확성이 높다는 것 즉 信賴度가 높다는 것을 일으켜 준다.

다음으로 精密度는 無作爲誤差에 關聯된 것으로서 反復測定에서 얼마나 再現性이 있느냐 하는 것과 關係가 있다. 例를 들어 說明하면 그림에서 보는바와 같이 신식총은 상대적으로 정확도는 낮으나 精密性은 높다고 할 수 있고 구식총은 정확도는 신식총보다 높으나 精密度는 낮다할 수 있다.

따라서 正確度는 精密度가 높아야 확보될 수 있는 것이지 精密度가 낮으면 正確度 또한 낮다. 즉, 精密度는 正確度의 必要條件인 것이다.

이렇게 正確度(accuracy)과 精密度(precision)을 強調하며 說明하는 理由는 精密性의 정도에 따라 신뢰도가 달라질 수 있기 때문이다. 가령 센티미터(cm)에는 신뢰도가 있으나 미리미터(mm)에는 그렇지 않은 尺度를 가지고 10cm 5mm, 10cm, 7mm의 두 測定對象을 측정할 경우는 신뢰도가 문제가 되는 것이다. 따라서 신뢰도는 주어진 정밀성의 정도와 관련되



〈正確度와 精密度間의 關係〉

이 이해되어져야 하며 요구되는 정밀성의 정도가 높으면 높을수록 적절한 신뢰도를 확보하기가 더 어렵게 되는 것이다. 즉 정밀성이 덜 필요할 때 고도의 신뢰도를 구하기가 상대적으로 낫다. 따라서 높다. 낮다와 같은 두 개의 값만을 추구하는 것이, 다섯개의 값을 추구할 때 보다 신뢰도가 높은 측정도구를 개발하는 것이 용이한 것이다.

그런데 科學이 발전됨에 따라 측정의 정밀성은 증가하는 傾向이 있으며, 이것은 科學者가 命題 특히 상관관계의 程度와 形態의 具體化를 보다 분명하게 측정하도록 해준다 예를 들어 과거에는 기후를 딥다, 춤다라는 용어를 사용하여 묘사했지만 지금은 온도계를 사용하여 묘사하는 진전을 보이고 있는 것이다. 이렇게 온도계를 사용함으로써 요구되는 정밀성이 정밀성이 높은 것에도 불구하고 높은 신뢰도를 확보할 수 있는 것이다. 그런데 여기서 주의 할 것은 신뢰도가 충분히 높지 않는 한 고도의 정밀성을 나타내게 하는 多重의 카테고리를 설정하는 것이 쓸모도 없을 뿐만 아니라 잘못을 유발할 수도 있다는 점이다. 가령 25암페어에서 27암페어사이를 웠다갔다하는 전류의 눈금을 26,000암페아로 읽어서는 안된다는 것이다. 이것은 오히려 26암페아나 아니면 26 ± 1 암페아로 읽어야 할 것이다.

2. 信賴度의 理論的 接近

우리는 앞에서 信賴度란 동일현상에 측정도구를 반복적용하여 동일한 결과를 얻게 되는 정도라고 정의할 수 있음을 보았는데, 이것은 다른 말로 하면 측정도구의 적용결과 오차가 발생하는 정도라고 정의 할 수도 있는 것이다. 그런데 이러한 誤差는 자(尺)나 저울같은 물리적인 測定道具를 사용할 때보다 질문지등의 측정도구를 사용할 때에 더 많이 나타나는데 이것들은 社會科學에서 사용되는 측정도구가 對象을 間接的으로 측정하기 때문이다. (Nahmias, 1976, p. 65)

우리는 이러한 誤差를 이용하여, 신뢰도에 대한 개념을 公式化할 수 있다. 이것은 신뢰도에 대한 概念定義를 보다 분명하게 해줄 뿐만 아니라 실제 신뢰도를 측정하는 것을 매우 쉽게 해준다. 다시 말해서 이것은 測定值를 統一性있게 설명해주며, 또한 다양한 分析技術

을 이해하는 데 기초가 될 것이다.

그리면 信賴度의 公式化는 어떻게 이루어지는 것인가? 개개의 測定(xi)은 두 가지 요소, 즉 진실된 값(ti)과 誤差(ei)로 구성되어 있으며, 신뢰도는 측정하려고 하는 값(xi)의 分散에 대한 진실된 분산의 비율로 정의할 수 있다. 이것을 수식으로 표현하면 다음과 같다.

$$xi = ti + ei \quad (4.1)$$

xi =어떤 측정대상 (i)에서 획득한 실제값.

ti =어떤 측정대상에 대한 진실된 값.

ei =어떤 측정대상에 발생하는 오차

이것을 分散으로 표현하면 다음과 같다.

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 \quad (4.2)$$

σ_x^2 =실제 값의 분산(全分散)

σ_t^2 =진실된 값의 分散.

σ_e^2 =오차의 分散

이 식은 몇 가지 사실을 암시해준다. 즉, 어떤 측정대상을 동일한 측정도구로 100번 측정한 것보다는 100,000번 또는 100,000번 측정한 것의 平均이 보다 진실에 가깝다는 것이다. 그러나 아무리 진실에 가깝다고 해도 이것이 全分散(σ_x^2)보다 클 수는 없다. 왜냐하면은 오차를 포함하고 있지만 x_i 에는 오차가 없는 값이기 때문이다. 그런데 현실적으로 특히 사회과학의 측정에서 오차가 없다는 것은 생각하기 어렵기 때문에. 이러한 誤差값을 계산하여 진실된 값을 구할 수 있을 것이다. 즉 誤差分散을 계산함으로써 信賴度를 측정할 수 있는 것이다.

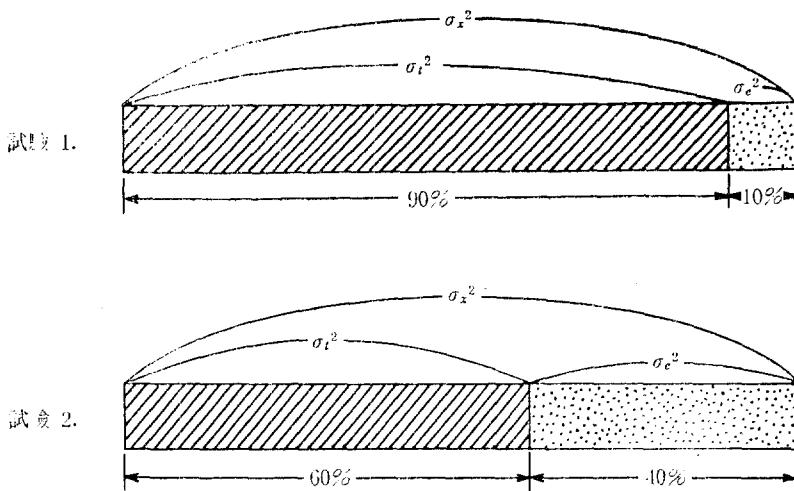
다시 말해서 실제 값의 分散(全分散)에 대한 진실된 값의 分散의 比率인 신뢰도는 다음과 같이 표현할 수 있는 것이다.

$$r_{xx}(\text{信賴度係數}) = \frac{\sigma_t^2}{\sigma_x^2} = \frac{\sigma_x^2 - \sigma_e^2}{\sigma_x^2} \quad (4.3)$$

여기서 측정이 전부 잘못된 것이라면 $\sigma_x^2 = \sigma_e^2$ 이 되어 신뢰도 계수는 0이 될 것이며, 오차가 없으면 $r_{xx} = \sigma_t^2$ 가 되어 신뢰도 계수는 1이 될 것이다. 이렇게 신뢰도 계수는 0에서 1사이를 움직이는데 1로 갈수록 신뢰도가 높은 것을 의미하고 0으로 갈수록 신뢰도가 낮은 것을 의미한다. 다른 말로 하면 誤差分散이 적으면 적을수록 신뢰도는 높아지는 것이고 많으면 많을수록 그 测定의 信賴度는 낮아지는 것이다.

이러한 관계를 그림으로 표시하면 보다 쉽게 이해할 수 있다. 〈圖 2〉 즉 「試驗 1」에서는 오차분산이 10%이고 「試驗 2」에서는 오차분산이 40%이므로 前者の 测定이 後者에 비해 훨씬 신뢰도가 높은 것이다.

圖 2. 信賴度係數의 圖示



3. 信頼度의 測定

信頼度를 측정하는 方法에는 여러가지가 있으나 일반적으로 많이 사용되는 것으로는 再検査法, 積數樣式法, 半分法等이 있다.

(1) 再検査法

再検査法(Test-retest Method)은 信頼度의 概念의 定義와 매우 밀접하게 연관되어 있으며, 보통 신뢰도를 측정하는 방법으로 가장 많이 사용된다. 이것은 동일한 대상에 동일한 측정도구를 서로 상이한 시간에 두번 측정한 다음 그 결과를 비교하는 것이다. 다시 말해서 어떤 퀴즈에서 측정을 한 후 일정시간이 경과하고 난 뒤 동일한 测定道具를 사용하여 동일한 응답자에게 다시 한 번 측정을 하여 그 결과의 상관관계를 계산하는 것이다. 여기서 도출된 相關係數(Correlation Coefficient)가 신뢰도의 정도를 추정해준다. 즉 상관계수가 높다는 것은 신뢰도가 높다는 것을 의미한다. 이렇게 볼 때 이 再検査法은 安定性을 強調하는 方法이라고 할 수 있다.

그러나 이 방법은 다음과 같은 한계가 있다. 첫째, 동일한 측정도구를 두번 적용함으로써 앞의 퀴즈가 두번째 측정에 영향을 미친다는 것이다. 예를 들어 측정도구로 질문지를 사용할 경우 처음 질문에 익숙해진 응답자는 두번째 질문에서는 처음 질문에서와 똑같은 형태로 응답할 가능성이 많은 것이다. 이것은 결국 信頼度를 실제보다 더 높게 추정되게 할 것이다.

이러한 영향은 時間의 간격에 크게 달려있다. 즉 檢査와 再検査 사이의 기간이 짧을수록 대상자의 기억이 생생하여 일관성있는 행동이 이루어질 것이며 이것은 신뢰도를 매우 높게 할 것이다. 반면 그 기간이 길면 길수록 신뢰도는 상대적으로 낮게 나타날 것이다.

때문에 신뢰도가 지나치게 높게 나타나는 (over-estimated) 것을 방지하기 위해서는 적용 시간간격을 넓히는 방법을 생각해볼 수 있다. 조사자에 따라 차이는 있지만 보통 합리적인 간격으로 1週에서 4週까지가 적당한 것으로 생각되고 있다(Black & Champion, 1976, p. 242).

그런데 여기서 첫번째의 測定道具의 適用이 두번째의 適用에 어느 정도 영향을 미치는가를 测定할 수 있다. 즉 對象者를 實驗集團과 統制集團으로 나눈 다음 실험집단에서는 척도를 두번 더 적용하고 통제집단에서는 한번만 적용하여 A' 과 A 의 차이가 A 와 B 의 차이보다 크면 척도의 첫번째 적용이 다음 적용시에 영향을 미쳤다고 할 수 있는 것이다. 그런데 이런한 影響이 크다고 하는 증거가 명백하면 이 檢查法에 의해 확보된 信賴度係數는 그대로 믿을 수가 없게 된다.

A	A'
	B

둘째, 外部變數의 영향을 파악하기 곤란하다는 점이다. 즉 測定道具의 적용구간사이에 측정속성이 어떤 變化가 일어날지 모른다는 것이다. 다시 말해서 처음과 두번째 측정에 있어 속성에 變化가 일어나면 조사자는 측정결과의 차이가 속성의 變化에 의한 것인지 信賴度가 낮아 그런 것인지 구별하기가 어려운 것이다. 이것은 신뢰도를 오히려 떨어뜨리는作用을 할 것이다. 이렇게 볼 때 再檢查法은 測定道具의 올바른 신뢰도를 過多測定할 수도 있고 過少測定할 수도 있다고 할 수 있다. 다시 말해서 再檢查法에 의해 고도의 신뢰도를 조사자가 확보할 경우에는 모든 것이 잘 되어 문제가 없다. 그러나 높은 신뢰도를 확보하지 못하는 경우는 조사자는 그 이유를 설명할 수가 없는 것이다. 따라서 이 방법은 外的要因이 작을 때 보다 유효하게 사용될 수 있다.

대신 이 방법의 장점은 측정도구 자체를 직접 비교할 수 있다든가, 적용이 간편하고 평가가 용이한 점을 들 수 있다.

(2) 複數樣式法(multiple forms technique)

이것은 平行樣式法(parallel-forms technique)이라고도 하는 것으로, 再檢查法의 두가지限界를 어느정도 극복하게 해 준다. 이것은 평행을 이루는 두가지 형태의 측정도구를 사용하여 각각 독일한 표본에 차례로 적용하여 봄으로써 신뢰도를 측정하는 방법이다. 예를 들어 集團結束力(group cohesion)에 관심이 있는 조사자는 다음과 같은 두가지 측면에서의 측정을 할 수 있을 것이다. 즉 하나는 어떤 집단이 개인에게 주는 매력이고 다른 하나는 그 집단의 구성원으로서 얻는 威信이다. 이 경우 測定值간의 상관관계가 크면 신뢰도가 높은 것이고 그렇지 않으면 낮은 것이다.

이러한 複數樣式法은 測定道具에 포함된 항목이나 문항들이 어떤 측정대상의 母集團의

한 표본에 지나지 않는다는 논리에 근거를 둔 것으로 다음과 같은 문제점이 있다. 즉 두 개의 양식을 평행하게 하는 것이 어렵다는 것이다. 다시 말해서 동일한 현상을 측정하는데 사용될 두 개의 同等한(equivalent) 测定道具를 개발하는 것이 어렵다는 것이다. 그리고 또 이 방법에 의해 확보된 신뢰도가 낮을 경우 이것이 측정결과가 본래부터 신뢰도가 낮아서 그런 것인지 아니면 두 개의 양식을 等同하게 하는데 실패한 때문에 그런 건지 설명할 수가 없다는 것이다.

그리고 또 측정양식을 차례로 만들어 두 양식간에 높은 反復性(repeatability)을 보일 경우 똑같은 體系的 誤差가 각 양식에 포함될 수 있다는 것이다. 예를 들어 두명의 관찰자에게 동일한 관찰기법을 사용하도록 훈련시킬 경우 관찰결과는 아마 높은 反復性(신뢰도)를 보일 것이다. 나, 체계적 오차를 피할 수 없다는 것이다. 다시 말해서 두개 양식의 测定道具가 같은 局域(continuum)을 测定할 수 있도록 相關關係가 높다면 被檢查者에 對해서도 이들의相互關係가 명백할 것이므로 처음 양식의 적용은 두번째 양식의 적용에 영향을 미칠 수 있다는 것이다.

(3) 半分法(Split-Half Method)

이 방법은 측정도구를 임의로 반으로 나누어 각각 독립된 두개의 척도로 사용함으로써 신뢰도를 측정하는 방법이다. 다시 말해서 調查項目의 반을 가지고 조사결과를 획득한 다음, 항목과 다른 반쪽을 동일한 대상에 적용하여 얻는 결과와 비교하는 방법이다. 이 방법은 複數樣式法을 논리적으로 보다 발전시킨 것으로서 項目이나 試驗이 여러가지여서 半分이 가능한 尺度에만 적용할 수 있다. 다시 말해서 半分法은 다음 두 가지 前提條件이 필요하다, 즉 ① 그 测定道具가 經驗的으로 單一의이라는 것이 明白해야 한다. ② 兩分된 各測定道具의 項目數는 그 自體가 各各 完全한 尺度를 이룰 수 있도록 充分히 많아야 한다. 半分된 項目數는 적어도 8개 내지 10개가 있어야 하며 따라서 全體項目數는 16 내지 20個의項目을 가지고 있어야 한다. (Goode & Hatt, 1952, p. 236) 그리고 半分하는 방법에는 보통項目을 짹수 항목과 홀수항목으로 반분하거나 아니면 無作爲的으로 항목의 반을 추출하여 나머지를 조 하나의 尺度로 만드는 식으로 일반적으로 사용되고 있다. 그런데 실제로는 두 부분을 따로 떼어 적용하는 것이 아니라 내용적으로만 반으로 갈라놓고 본래의 측정도구를 그대로 사용한다. 그리하여 두 부분의 측정결과를 비교하여 상관관계를 계산함으로서 신뢰도를 측정하는 것이다. 이렇게 하여 얻어진 상관계수를 시정하기 위해서 Spearman-Brown 공식을 적용하게 된다.

$$r_{xx'}' = \frac{2r_{oe}}{1+r_{oe}} \quad (4.4)$$

r_{xx}' ; 본래의 신뢰도계수

r_{oe} ; 반분법으로 얻은 신뢰도계수(reliability coefficient)

이것은 질문의 수가 짹수 개인 질문지가 홀수 개인 질문지보다 신뢰도가 높고 또 질문지 전체가 반수 보다 신뢰도가 높다는 것을 전제로 한다.

이半分法은 項目的 同質性을 의미하는 内的一貫性(internal consistency)을 강조하는 방법으로 어떤 특정项目的 신뢰도를 정확히 파악할 수 없는 단점이 있다.

이제까지 설명한 測定道具에 대한 信頼度의 測定은 이것을 조사 시작전에 하는 것이 좋다. 測定道具에 오차가 있다면 意味있는 결과를 달성할 가능성이 줄어들기 때문에, 신뢰할 수 없는 측정도구를 가지고 조사를 계속하는 것보다는 오히려 조사를 연기하여 신뢰도를 제고시키도록 하는 것이 현명한 것이다. 다시 말해서 調査는 각段階에 따라 진행되어 절야하며 타당성 있고 신뢰할만한 측정도구의 개발이 항상 어떤 현상에 대한 조사보다 선행되어야 한나는 것이다.

그러나 실제로는 조사가 단계적으로 이루어지기 보다 오히려 앞뒤로 진행되는 것 같다. 즉 측정도구가 불완전하고 심지어 매우 조잡하다고 할지라도 실제로 假說檢定에서 效用化 할 때까지 이것을 개선한다는 것이 그리 쉬운 일이 아닌 것이다. 따라서 종종 측정도구를 조잡한 채로 그대로 사용하는 것이 오히려 더 유리할 때도 있다. 이것은 특히 그러한 측정도구를 가지고 실제 조사를 해보아야 改善點을 발견할 수 있을 때 더욱 그러하다.

4. 信頼度의 提高方法

信頼度를 높이는 방법은 진실로 값의 分散, 즉 개개 측정사례의 차이에 대한 分散을 극대화하는 即便 誤差分散을 극소화하는 것이다. (式 4.3 참조) 이것을 maxmincon 原理라고 하는데 이 원리를 실현하는 일반적인 절차는 다음과 같다.

첫째, 測定의 실패가 概念에도 그 원인이 있으므로 測定道具가 되는 項目을 누구에게나 동일하게 이해되도록 명백하게 구성해야 한다. 애매모호한 항목은 측정대상이 이것을 서로 다르게 해석할 수 있기 때문에 誤差分散이介入할 여지가 생기게 되는 것이다. 이러한解釋의任意性는 3차분산을 크게 하여 결국 신뢰도를 떨어뜨리는 결과를 가져오게 되는 것이다. 둘째, 어떠한 요인이 측정의 신뢰도를 떨어뜨리는가를 결정하기 위하여 測定狀況自體에 대한 分析는 하도록 한다. 표준화되고 잘統制되어 또類似한 條件下에서 적용되도록 해야 한다. 측정도구의 적용에 있어 그 條件이 달라지면 다시 誤差分散이介入할 가능성이 있기 때문이다.

셋째, 심리적 측정과 관련되어 개발된 방법으로 처음에 사용된 것과 같은 형태의 측정도구를 추가로 더 사용하도록 하는 것이다. 다시 말해서 측정도구가 충분히 밀을만한 것이 못될 경우 동일한 종류와 질을 가진 項目을 추가로 더 사용한다는 것이다. 項目數를 보다 많아한다는 것은 實際 測定值(xi)가 진실된 값(ti)에 보다 근접될 가능성이 높다는 것을 의미하며 이것이 곧 신뢰도를 증가시키는 것이다. 이는 마치 標本抽出을 할 때 표본의 數를 많아함으로써 우연에 의한 오차를 적게 하는 원리와 같다.

그러나 項目數를 많이 한다는 것은 실제 측정된 값의 구체적 내용을 파악하는 것을 그만큼 어렵게 하며, 따라서 항목수의 증가가 신뢰도를 높여주기는 하나 理論의 開發에 사용하기에는 기나치게 이질적이라는 결과를 초래하게 될 수도 있다는 點을 명심할 必要가 있다. 넷째, 測定道具가 되는 각 項目的 성격을 비교하여 서로 對照的인 項目들을 比較分析하는 방법이다. 이것은 項目的 性格이 서로 對照的인 것은 그 結果도 서로 대조적으로 나타나야 한라는 논리에서 개발된 것이다. 여기에서는 相關係數가 零에 가까워지는 것이 신뢰도가 높다는 것을 의미한다.

다섯째, 측정도구의 사용이나 응답에 있어서 가능한 한 분명하고 표준화된 지시나 설명을 함으로써 측정오차를 줄이도록 해야 한다. 불분명한 지시나 설명은 오차분산을 크게 할 것이기 때문이다.

Cronbach, Rajartnam, 및 Gleser는 이제까지 설명한 信賴度의 傳統的 概念에 대한 改定을 시도하고 있다. 즉 이들은 信賴度理論의 主要한 關心을 測定結果가 測定하려는 對象의 母集團을 어느정도 代表할 수 있는가에 있기 때문에 信賴度 대신 一般化可能性(generalizability)의 概念을 도입해서 사용해야 한다는 것이다. 여기서 一般化可能性이란 우리가 일련의 測定에서 실제로 알고 싶은 것이 무엇인가 하는 것을 의미한다. 즉, 어떤 測定結果가 우리가 적용할 수 있는 測定對象의 潛在的 母集團에서 채택된 다른 어떤 측정결과와 어느정도 유사하며 또 어느정도 차이가 있는가 하는 것이다. 이러한 의문이 바로 一般化可能性을 의미하는데 그러한 測定結果間의 特定關係가 신뢰도 또한 일반화가능성이 있는 증거로 볼 수 있느냐 없느냐 하는 것은 주변條件와 測定의 유사성과 차이점을 어떻게 규정하는가에 달려있다. 이러한 測定에서 무엇이 동일하고 무엇이 서로 다른가를 구성하는 것은 調查研究가 알고자 하는 것이 무엇인가에 달려있다.

第3節 妥當度와 信賴度의 相互關係

測定이 있어서 妥當度와 信賴度의 重要性은 더 말할 나위가 없으나 그 중요성의 촛점은 妥當度가 보다 기울어져 있다. 測定評價의 主要한 關心이 조사자가 측정하고자 하는 바를 얼마나 정확하게 측정했느냐 하는 관점에서부터 시작되었다고 볼 때 역시 신뢰도보다 타당도의 意味가 더 크다 하지 않을 수 없다.

다시 말해서 測定道具의 信賴度가 타당도에 비해 확보하기가 용이한 이점이 있긴 하지만 일반적으로 타당도의 수립이 보다 더 가치가 있는 것이다. 예를 들어 어떤 측정도구가 동일한 현상에 대해 반복 적용되어 동일한 결과를 얻었다고 하면 이 측정도구는 전적으로 신뢰할 수 있다고 할 수 있다. 그러나 이것이 측정을 정확하게 했다는 것을 의미하지는 않는다. 즉 신뢰도가 완전한 측정도구라 하더라도 타당도가 낮은 것은 있을 수 있지만 타당도가 완전한 측정도구는 신뢰도도 완전한 것이다. 따라서 사회과학에서 사용하는 측정도구가

충분히 타당도가 있으면 신뢰도는 방법론에서 그 의미가 없다고 할 수 있다. 그러나 실제로는 많은 경우 타당도가 결여되어 있으므로 신뢰도의 가치가 있는 것이다.

이러한 妥當度와 信賴度의 關係를 보다 구체적으로 살펴보면 다음과 같다.

첫째, 妥當性 있는 測定은 항상 信賴性이 있다.

둘째, 妥當性이 없는 측정은 신뢰도가 있을 수도 있고 없을 수도 있다.

셋째, 信賴性이 있는 測定은 타당도가 있을 수도 있고 없을 수도 있다.

넷째, 신뢰성이 없는 측정은 타당도가 없다.

이러한 信賴度와 妥當度의 關係를 Lord F.M.와 Novick M.K.은 다음과 같은 式으로 표시하고 있다.

$$V \leq \sqrt{k} \quad (v; \text{타당도}, k; \text{신뢰도})$$

第4節 妥當度와 信賴度에 영향을 미치는 要因

妥當度와 信賴度에 영향을 미치는 요인에는 여러가지가 있으나 여기서는 測定道具의 効用性에 영향을 미칠만큼 중요한 것 몇가지만 설명하기로 한다. 이를 要因은 크게 4가지로 나누어 생각해 볼 수 있는데, (1) 檢查道具 및 그 内容, (2) 環境的 要因, (3) 個人的 要因, (4) 調査者의 解釋 等이 그것이다.

1. 檢查道具 및 그 内容

측정의 신뢰도와 타당도를 평가할 때 제일 먼저 해야 할 일은 측정에 쓴 項目이 전부 포함되어 있느냐 하는 것이다. 또 각 項目은 측정하려고 하는 現象의 屬性이나 特性을 잘 대표하고 있는지를 알아보는 것이다. 조사자가 不注意 또는 실수로 항목을 누락한다든지 하면 타당도를 저해하게 되는 것이다. 이러한 요인을 몇 가지 살펴보면 다음과 같다.

(1) 測定의 길이.

測定이 짧어서 조사대상자가 짐짓을 느끼게 되면 이들은 실제 느낀점을 이야기하기보다 편의 위주로 응답할 가능성이 있다. 따라서 조사자는 응답이 형식적으로 이루어진 것이 없는지를 세심하게 살펴보아야 한다. 形式的 應答이란 예를 들면 절대 찬성에서 절대 반대까지로 된 리커트척도로 이루어진 조사항목의 응답에서 모두 절대 찬성에 응답하는 것과 같은 것을 말한다('물론 각 項目間에는 서로 대조적인 것이 있다고 가정한다') 말하자면 질문지의 길이가 길면 길수록 形式的應答(set response)이 발생할 가능성이 그 만큼 커지는 것이다.

(2) 文化的 要因

관례상 사용되지 않는 단어나 귀절을 포함하는 혹은 신뢰도가 떨어지며 따라서 타당도도 문제가 된다. 예를 들어 냉장고에 대해 아이스박스라는 용어를 사용하면 조사자가 의도한 의미와 조사 대상자가 이해하는 의미사이에는 오해가 생기게 될 것이다. 따라서 측정이 이

루어지고 있는 당시의 문화는 態度測定의 신뢰도와 타당도를 확보에 영향을 주는 중요한 요인으로 작용한다.

(3) 開放式質問과 閉鎖式質問

개방형 질문에서는 응답자가 직접 자기 표현을 하는 것이므로 응답자의 능력, 教育水準 등이 타당도와 신뢰도에 영향을 미친다.

(4) 機械的인 要因

여기서 기계적 요인이라란 질문지에 있어서 탈자, 오자, 읽기 어려운 단어, 페이지의 누락 등과 같은 것을 말하는데 이것은 오해를 유발하여 측정의 타당도와 신뢰도를 저하시킨다. 이것은 對人面接에서도 나타나는 것으로, 즉 면접자가 질문지의 질문 이외의 설명을 한다든가 임의로 다른 단어로 대체하여 일은 면접결과는 신뢰도를 저하시키는 것이다. 따라서 이상적으로는 각 응답자에 대한 측정도구의 적용조건을 동일하게 적용하도록 해야 할 것이다.

2. 環境的 要因

측정이 적용되는 환경적 요인 또한 측정의 타당도와 신뢰도에 영향을 주는 중요한 요인이라고 할 수 있다. 환경적 요인 중 특히 중요한 것만 몇 가지 살펴보면 다음과 같다.

(1) 對人面接과 自己記述式質問紙(self-administered questionnaires)

어떤 정대상에게 똑같은 질문을 하여도 그 것이 직접 면접자와 얼굴을 대하는 면접인 경우와 대상자가 직접 문항을 기재하는 경우는 상당한 태도상의 차이가 있다. (Boruch, 1971, pp. 308-311)

(2) 측정도구를 完成하는데 필요한 지시의 明白性

조사대상자가 측정을 받는데 필요한 절차를 분명히 하지 않으면, 무의식중에 정보를 잘 못 제공할 가능성이 많은 것이다.

3. 個人的 要因

측정의 신뢰도와 타당도에 영향을 미치는 개인적 特性 중 중요한 것만 몇 가지 살펴보면 다음과 같다.

(1) 應答者의 社會經濟的 地位

社會經濟的 地位의 일차적 構成要素로는 職業, 教育水準, 所得, 倫理的 背景等을 들 수 있다. 이 한 要因은 成果測定 및 態度調查의 差異를 유발하게 되므로 조사자는 조사하려는 대상의 사회경제적 수준에 되도록 상응하도록 측정도구를 작성할 필요가 있다.

(2) 年齢, 性別 및 成熟度

연령 및 성별은 人間의 成熟水準과 관계 되는데 이것은 調查에 대한 이해에 영향을 미친다.

(3) 記憶力

검사를 間後에 걸쳐 두번 실시할 때 처음 검사에서의 응답을 기억하는 능력은 두번째 검사의 응답에 영향을 미친다.

(4) 社會的要請(social desirability)

사회적으로 바람직한 것으로 인정되는 것이 포함될 경우 이것은 응답자의 전반적인 응답 형태에 중요한 영향을 미친다. 즉 응답자들은 자신들의 진실을 밝히기보다 사회적으로 바람직하다고 하는 것을 응답할 가능성이 많다는 것이다. 확실히, 측정의 타당도와 신뢰도에 영향을 미치는 이러한 사회적으로 바람직한 要請의 効果를 무시해서는 안될것이다. (Larsen, 1958)

4. 調査者の解釋

調査者が 結果를 어떻게 해석하는가에 따라 신뢰도와 타당도에 차이가 있다. 즉 調査者は 자신이 수집한 정보를 자기가 원하는 대로 符號化할 수 있는데, 신뢰도와 타당도를 높히기 위해서는 符號化의 客觀性을 最大化해야 한다. 그리고 어떤 태도의 測定 결과 획득한 점수의 해석이 신뢰도와 타당도에 영향을 미치는 것이다.

[要約]

1. 타당도는 측정하려고 하는 것을 어느 정도 측정했는가를 파악하는 것이다.
2. 타당도의 형태에는 세가지, 즉 論理的 타당도, 경험적 타당도, 구조적 타당도가 있는데, 이들은 妥當度를 평가하는 수단으로서 論理的 檢證, 經驗的 檢證, 論理的 및 經驗的 檢證을 각각 사용한다.
3. 구조적 타당화 방법에는 論理的 過程, 수렴차별개념 또는 多重屬性 多重方法行列技法 실제 調査에 시의 例, 要因分析에 의한 方法等이 있다.
4. 測定의 妥當化方法에는 論理的 妥當化, 專門家意見, 既知集團法, 獨立基準法이 있는데 이들을 동시에 사용하거나, 論理的 妥當化를 중심으로併用하는 것이 좋다.
5. 信賴度는 동일현상을 동일 측정도구로 반복적용하여 동일한 결과가 나오는 정도를 말한다.
6. 신뢰도는 타당도 보다 확보가 용이하고 측정도구의 완전여부를 가려준다.
7. 신뢰도의 개념은 安定性 正確性無作爲誤差라는 세 가지 측면에서 고찰되는데 최근에는 一般化可能性的 概念으로 바뀌고 있다.
8. 신뢰도를 이론적으로 설명하면 全分散에서 진실된 값의 分散比率로서 이 비율을 신뢰도 계수라고 하는데 신뢰도 계수가 높으면 신뢰도가 높다.
9. 신뢰도를 측정하는 방법에는 再検査法, 複數樣式法, 半分法等이 있으며 이들은 각각 장단점이 있다.

10. 신념도를 높히려면 概念의 明白化, 測定狀況分析, 測定道具項目의 追加, 項目的 比較分析 표준화된 지시 또는 설명을 하는 것이 좋다.
11. 신념도와 타당도의 관계는 $V \leq \sqrt{k}$ 식으로 표현된다.
12. 신념도와 타당도에 영향을 미치는 요인으로 檢查 및 그 內容, 환경적 요인, 개인적 요인, 조사자의 해석을 들 수 있다.