Content and Cultural Validity of Scales Within the Korean Context

Hana Song

Department of Child Psychology and Education, Sungkyunkwan University, Seoul, Korea

연구도구의 내용 타당도와 문화 타당도

송하나

성균관대학교 아동청소년학과

이번 호에는 아동학 연구자들이 사용해 온 연구 도구들의 이론적 적합성과 심리측정 이슈에 대한 리뷰가 게재되었다. 이 논평의 저자인 나 자신을 포함하여 많은 연구자들이 연구 도구 개발과 사용에 대해 간과했던 부분들을 다시 생각하는 계기가 될 것이다.

어떤 변인을 측정하기 위한 도구를 개발(development) 한다는 것은 사전적으로 ‘개척’의 의미를 담고 있다. 즉, 연구자가 문항을 새롭게 생성하여 기존의 도구와는 차별화된 새로운 도구를 만드는 것을 말한다. 문항을 생성하는 방법은 다양하다. 그러나 학술적 의의를 가진 좋은 도구를 개발하기 위해서는 먼저 선행 연구 탐색을 통해 측정하려는 변인의 속성과 구조를 이론적으로 철저히 이해하는 과정을 거쳐야 한다. 검사를 개발하는 첫 단계에서, 변인의 속성이 무엇인지, 어떠한 요인 구조로 정의되어야 하는지, 어떤 하위 요인을 가정해야 하는지 등을 생각하고 이에 따라 문항의 내용을 결정하는 것이 적절하다.

무엇보다 문항의 내용이 그 변인의 속성을 제대로 타당하게 측정하고 있는지 검증하는 것은 중요하다고 생각된다. 내용 타당도 확보를 위해 문항을 만들기 전에 심층 인터뷰나 개방형 설문을 해보는 것은 유용하다. 전문가 집단과 함께 문항의 적합성을 토론하는 것 역시 필수적이며, 관련 지식이 없는 일반인들에게 유사한 문항들끼리 분류하도록 하는 간

단한 과제로도 타당도를 대략 알아볼 수 있다. 좋은 도구일수록 여러 학자들이 오랜 기간 협업하여 논의하는 과정을 거친다. 그 예로, 지능검사는 1905년 처음 만들어진 이후 112년 동안 수많은 연구를 통해 개정되어 왔다.

하지만, 충분한 이론적 논의와 검토 없이 기존 검사 문항의 문구나 문장을 조금 수정한다든지, 유사한 검사 도구에 있는 문항들만으로 설문지를 구성한 뒤 요인 분석을 실시하여 최종 문항을 구성한다면, 이것을 도구의 개발이라고 할 수 있을지 의문이다. 새로운 도구 개발이 아니라, 검사 도구의 타당도 검증 연구가 아닌가 한다.

물론 도구를 개발할 때 기존 문항들도 사용이 가능하다. 그러나 사용된 문항들의 출처를 구체적으로 밝히고 이론적 근거를 설명해야 한다. 특히, 외국 학자들이 개발한 연구 도구를 번안해서 사용하는 경우에는 내용 타당도 뿐 아니라 문화적 타당도를 함께 고려해야 할 것이다. 연구자들은 원 도구와 번역된 도구의 측정 기준을 동일하게 맞추기 위해 번역과 역번역을 수행하고 문항반응이론을 적용하여 동질성을 분석한다. 이처럼 번역된 도구의 문화적 동질성을 확보하는 것은 외국인을 준거로 만들어진 도구를 우리나라에 적용하거나 비교문화 연구를 수행하는데 중요한 전제조건일 수 있다.

하지만 문화적 동질성을 문화적 타당성과 동일한 것으로

생각해서는 안된다. 오래 전 아들을 출산한 미국 친구와 동양의 남아선호사상에 대해 이야기를 나눈 적이 있다. 그녀가 “아들은 딸보다 무겁고 힘도 세서 키우기가 더 힘들어. 가족들의 도움이 많이 필요해. 시어머니도 그래서 아들을 낳으면 며느리를 더 격려하고 잘해 주시는게 아닐까?” 하고 말한 순간만큼 문화적 차이를 크게 느꼈던 적이 없는 것 같다.

이처럼 외국 연구자들과 이야기를 나누다 보면 어떤 현상에 대한 해석의 차이를 실감하는 경우가 있다. 나 자신 역시 번안된 설문지를 종종 사용하고 있음에도 불구하고, 어떤 변인을 우리 문화적 관점에서 완전히 새롭게 정의할 필요성을 종종 느낀다. 한국 아동과 청소년의 자존감은 어떤 속성을 가지고 있는지, 어떤 문항이 한국 부모들의 양육 태도를 가장 잘 나타낼 수 있는지 알아보기 위해 현상학적 혹은 문화기술적 연구들이 더 많이 수행되었으면 하는 바람이다.

양적 연구이든 질적 연구이든 내용과 해석의 타당성은 좋은 연구의 필수적인 조건이다. 그러나 번안된 도구를 사용하는 경우에 신뢰도는 충실히 보고하는 반면, 타당도에 대한 구체적인 논의를 간과하는 일이 상대적으로 많은 것 같다. 물론 여러 연구자들이 번역된 도구를 맥락에 맞게 수정하여 문화적 동질성을 검증하고, 준거 타당도나 구성 타당도를 제시하고 있다. 그럼에도 불구하고, 번안된 문항이 한국인들에게 적용가능한 것인지 아닌지를 생각하는 것 보다, 우리 고유의 사회문화적 관점에서 새로운 패러다임으로 변인을 정의하려는 노력이 있어야 하지 않을까 한다.

연구 도구의 질과 타당성은 통계분석으로 나타난 결과에 의해서만 결정되는 것이 아니라 문항생성 단계에서부터 내용의 적합성에 대한 심도 깊은 논의에 의해 더 크게 영향 받는다. 특히, 새로운 도구의 개발을 목적으로 한 연구를 수행하고자 하는 연구자들이라면 이러한 부분을 충분히 고려하여 문항생성 단계에 더 많은 노력을 기울였으면 한다.

In this issue of the *Korean Journal of Child Studies*, a review of theoretical adequacy and psychological measurement issues of the scales used by members of our academic society has been published. Many researchers, including myself as the author of this editorial, have had the opportunity to think about what they have overlooked regarding scale development and use.

The term *development* in scale development literally

means *pioneering*. In other words, scale development refers to the process of researchers generating new items and developing a new measure different from the existing ones. There are a variety of ways to generate items for a scale. However, in order to develop a good scale with academic significance, it is necessary to include the process of property and construct analyses of a variable through a deep examination of previous research. In the first step of scale development, researchers have to fully consider the attributes of the variables, the factor structures that should be defined for a variable, and the sub-factors that should be assumed, before they determine the contents of the items that should be listed.

The item generation procedure often includes conducting in-depth or focus group interviews and administering open-ended survey questions. Sorting tasks, which require people to classify items into similar categories, can also be used, and discussing the suitability and appropriateness of each item with other scholars is a good way to support content validity. Many scholars have collaborated for a long time to make good scales. For example, numerous studies to continuously revise intelligence tests have been carried out for nearly 112 years since the development of the first intelligence test in 1905.

However, if we construct a scale with only the items of existing scales, modify the items without sufficient theoretical review, and conduct factor analysis with the same items, it cannot be termed scale development. Rather, it would be called a study to examine the validity of a scale.

The use of existing items is possible when developing a scale, but the source of the items and theoretical background need to be clarified. In particular, when translating and using research scales developed by foreign scholars, cultural/ecological validity as well as content validity need to be considered. Many researchers carry out translation and back-translation procedures, and analyze the cultural equivalence between the original scale and a translated version of the scale using the item response models. This is an important prerequisite for conducting

cross-cultural studies.

However, cultural validity is not equal to cultural equivalence. A long time ago, I discussed the decline in the preference for sons in Asian culture with an American friend who had given birth to a son. I felt the cultural difference deeply when she said, “I think that it is harder to raise a son than a daughter, because boys are heavier and more powerful. I need a lot of help from my family. ... So, a mother-in-law provides more support and encouragement to the daughter-in-law when she has given birth to a son.”

In a similar way, I often realize cultural differences in the interpretation of a phenomenon when I converse with foreign researchers. Although I have used the translated versions of many scales, I was often required to acknowledge the need to define a variable in a completely different way within the Korean context. It may be useful to conduct a phenomenological or ethnographic study in order to find out the properties of self-esteem Korean children and adolescents possess, the items that can best represent the parenting attitude of Korean parents, and so on.

It is important to support the validity of the content and interpretation of a scale in both quantitative and qualitative studies. Surely, many researchers have tested the cultural equivalence and verified the validity and reliability of the translated versions of scales. However,

rather than merely verifying cross-cultural validity through an examination of whether the translated items are applicable to Koreans, there should be an effort to establish new theoretical frameworks to explain phenomena from our own sociocultural perspective.

The quality and validity of research scales are determined not by the results of the statistical analyses, but also by in-depth discussions of the validity of the items from the item creation phase. I hope that researchers involved in scale development studies take this into account and make more efforts to generate valid items.

Hana Song, Editor-in-Chief
of the Korean Journal of Child Studies.

Conflict of Interest

No potential conflict of interest relevant to this editorial was reported.

ORCID

Hana Song   <http://orcid.org/0000-0002-7745-7303>