

한국의 도구개발 간호연구에서의 타당도에 대한 고찰

이경희¹ · 신수진²

¹이화여자대학교 건강과학대학 간호과학부, ²순천향대학교 의과대학 간호학과

Validity of Instrument Development Research in Korean Nursing Research

Lee, Kyunghee¹ · Shin, Sujin²

¹Division of Nursing Science, College of Health Science, Ewha Woman's University, Seoul

²Department of Nursing, College of Medicine, Soonchunhyang University, Cheonan, Korea

Purpose: This integrative review study was done to analyze methods used for validation studies in Korean nursing research. **Methods:** In this study, the literature on instrument development in nursing research from Research Information Sharing Service (RISS) and major nursing journal databases in Korea were examined. The MeSH search terms included 'nursing', 'instrument', 'instrument development', 'validation' and 189 articles were included in the review. **Results:** The most frequently reported validity type was content validity, followed by construct validity, and criterion validity. One third reported a single type of validity, and 15% of the studies demonstrated three kinds of validity at the same time. In about 40% of the studies, both content and construct validity were examined. **Conclusion:** The results of the study indicate that it is necessary to provide a wider variety of evidence to establish whether instruments are valid enough to use in nursing research.

Key words: Validation studies, Validity, Instrument, Nursing

서론

1. 연구의 필요성

사물의 성질을 구체화하기 위하여 수를 부여하는 절차인 측정(measurement)은 다양한 사회과학 연구에서 중요한 관심사이며(DeVellis, 2003), 다른 학문에서와 마찬가지로 측정은 간호학 연구에서 핵심적인 역할을 한다(Park, 2005). 측정의 대상은 직접 측정이 가능한 것과 직접 측정이 불가능하여 간접 측정만이 가능한 대상으로 구분하며(Seong, 2002), 간호학에서 주요 연구개념인 스트레

스, 우울, 삶의 질, 자신감 등의 개념은 주로 추상적이고 주관적이며, 잠재적인 특성이기 때문에 직접적으로 측정하기가 어렵다. 따라서, 신뢰할 수 있고 타당한 도구를 이용한 측정이 필수적이다. 따라서, 간호학 연구에서 측정을 위해 도구의 신뢰도와 타당도 확보는 가장 중요한 요소이다.

신뢰도란 피험자들에게 동일한 검사를 반복 실시하였을 때 측정의 일관성(consistency)이라고 정의하고 있다(American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). 즉, 측정하고자하는 내용을 얼마나 정확하게 오차없이 측정

주요어: 타당화 연구, 타당도, 도구, 간호

*본 연구는 순천향대학교 학술연구비지원으로 수행하였음.

*This study was supported by the Soonchunhyang University Research Fund.

Address reprint requests to : Shin, Sujin

Department of Nursing, College of Medicine, Soonchunhyang University, 31 Soonchunhyang 6th Rd., Dongnam-gu, Cheonan 330-100, Korea

Tel: +82-41-570-2491 Fax: +82-41-574-3860 E-mail: ssj1119@sch.ac.kr

투고일: 2013년 10월 1일 심사외뢰일: 2013년 10월 21일 게재확정일: 2013년 11월 15일

This is an Open Access article distributed under the terms of the Creative Commons Attribution NoDerivs License. (<http://creativecommons.org/licenses/by-nd/4.0>) If the original work is properly cited and retained without any modification or reproduction, it can be used and re-distributed in any format and medium.

하였는지에 관한 검사도구의 일관성을 말하며, Cronbach's α 를 이용한 내적일관성 신뢰도와 Pearson 상관계수 추정 공식에 의한 검사-재검사 신뢰도가 일반적으로 보고되고 있다. 한편, 도구의 타당도와 신뢰도가 다른 개념이라 할지라도 분리하기 보다 연관성을 두고 분석하는 것이 두 개념을 이해하는데 도움이 되며, 타당도가 높기 위해서는 신뢰도가 높아야 하지만 신뢰도가 높다고 반드시 타당도가 높은 것은 아니다(Seong, 2002). 예를 들면 신뢰도가 높은 우울 측정도구로 불안을 측정할 경우 타당성이 떨어진다고 볼 수 있듯이 신뢰도가 높은 도구로 측정하였다고 해서 타당도가 반드시 높다고 볼 수는 없다. 따라서, 도구의 타당성을 증명하는 과정 또한 필수적인 과정이지만, 증명된 공식에 의해 수치를 산출하는 비교적 일반화된 신뢰도 검증과정에 비해 타당도를 증명하는 과정은 상대적으로 매우 다양한 접근을 통해 이루어지고 있다.

타당도는 피험자 집단에 사용된 측정도구나 검사에 의하여 얻어진 검사 결과의 해석에 대한 적합성이 검사자체의 문제가 아니다. 따라서, 편의상 타당도라고 표현하지만 정확히 말하면 해석에 대한 타당성을 말하는 것(Gronlund & Linn, 1990)이라는 주장을 시작으로 도구 사용을 통해 얻어진 해석에 대해 근거나 이론이 지지하는 정도라고 정의되고 있다(AERA et al., 1999). 즉, 개발된 도구가 의도한 바를 측정하고 있는지와 측정하고자 하는 것을 얼마나 충실하게 측정하였는가에 대한 것을 의미한다. 따라서, 타당도는 '있다 없다'라고 말하기 보다 도구를 이용하여 측정한 결과에 대한 해석의 적합성이 어느 정도인지를 의미한다는 것이다.

이상의 주장을 바탕으로 한다면 타당도는 근거를 통해 도구 사용 및 해석의 타당성을 증명하는 추론의 과정이므로 논리적이고 다양한 근거를 제시하여 추론의 적합성을 제시하고 도구의 타당성을 증명할 필요가 있다. 이는 타당도가 가능한 모든 증거들을 이론적 틀 안에서 종합하여 얻어지는 평가적 판단이라는 견해(Lee, 2008)와도 일맥상통한다. 이러한 이유로 타당도에 대한 접근 방법이 다양할 수밖에 없으며, 연구마다 다양한 용어와 다양한 분석방법이 적용되고 있어 독자에게 혼란을 야기하기도 한다.

지금까지 국내외적으로 간호학분야에서 많은 도구가 개발되어 사용되고 있음에도 불구하고 도구개발 시 타당도 연구가 어떻게 이루어지고 있는지에 대한 동향 파악이 매우 미진한 실정이다. 주로 개발된 도구에 대한 타당도 연구 동향은 청소년 우울이나 소아통증같이 특정한 대상에서 한 가지 개념을 측정하는 다양한 도구들을 대상으로 주로 어떠한 타당도 검증들이 이루어지고 있는지와 같이 제한된 범위에서 동향파악을 하고 있으며(Carnevale, 2011; Foster & Park, 2012), 영문으로 출간된 임상 연구를 대상으로 간호학분야에서 타당도의 경향연구(Ribeiro, Vedovato, de Moraes Lopes, Monteiro, & de Brito Guirardello, 2013)가 극히 제한적이었다. 이러한

제한점 때문에 국내의 간호학분야에서 도구 개발 연구의 타당도동향에 대한 선행 연구가 없어 본 연구에서 파악하고자 한다.

2. 연구 목적

본 연구의 목적은 국내 도구개발 연구동향을 분석하고 이들 중 도구의 타당화 과정을 체계적으로 고찰하여 그 특성을 파악하고, 추후 연구를 위한 방향을 제시하고자 함이며, 그 구체적인 목표는 다음과 같다.

첫째, 도구 개발 간호연구의 연구 및 도구관련 변인(출판년도, 도구적용대상, 간호개념, 연구유형)을 분석한다.

둘째, 도구 개발 간호연구의 타당도관련 변인(내용타당도, 구성타당도, 준거타당도, 기타)을 분석한다.

연구 방법

1. 연구 설계

본 연구는 도구 타당화에 대한 간호연구 동향 및 타당도 관련 자료 분석 및 결과 보고 동향에 대한 문헌 고찰 연구이다.

2. 연구 대상 및 자료 수집

본 연구는 간호연구 중 도구 개발 및 타당화와 관련된 국내 학술지 게재논문 중 특히, 2013년 6월 현재 학술진흥재단 등재지인 학술지에 게재된 논문을 선정 대상으로 하였다. 분석 논문은 초기 간호연구부터 2013년 6월까지의 연구로 한국교육학술정보원(www.riss.kr)과 간호관련 학술진흥재단 등재학술지의 홈페이지 데이터베이스를 검색하여 수집하였다. 검색어는 '간호(nursing)', '도구(instrument)', '도구 개발(instrument development)', '타당화(validation)'이었다. 국내 학술지는 모두 361편의 논문이 검색되었다. 일차로 검색된 논문의 제목과 초록을 확인하여 중복 논문 55개, 도구 개발 연구가 아닌 논문 117개를 제외하고, 총 189편을 표본으로 선정하였다(Figure 1).

3. 자료 분석 방법

연구 대상으로 선정된 189편의 논문에 대해 연구의 연구 및 도구 관련 변인, 타당도관련 변인, 타당도 근거 제시를 위한 분석 방법 및 결과 보고 동향을 분석하였다. 연구 및 도구관련 변인으로는 출판년도, 연구유형, 도구적용대상, 간호개념을 분석하였다. 타당도관련 변인으로는 내용타당도, 구성타당도, 준거타당도, 기타 타당화 근거

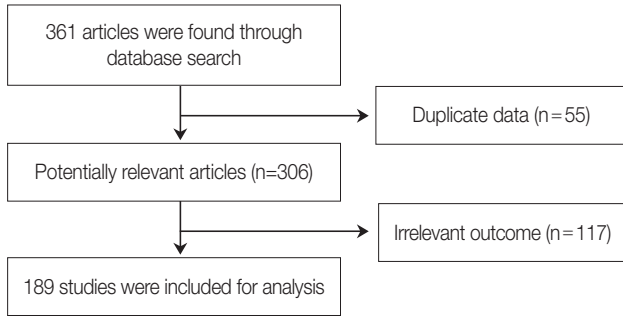


Figure 1. Article selection process flowchart.

에 대해 분석하였으며, 타당도의 유형에 대해서는 연구자가 사용한 용어를 그대로 적용하여 분석하였다. 타당도 근거 제시를 위한 분석 방법 및 결과 보고 동향 분석에서는 타당화 근거를 제시하기 위해 사용한 분석 방법 및 해석의 근거, 결과 보고 동향을 분석하였다. 간호개념분류를 위해서는 간호의 개념을 포괄적으로 포함하여 제시하고 있는 Nursing Outcome Classification (Choi et al., 2012)을 적용하였으며, 자료 분석은 SPSS 20.0 프로그램을 사용하여 빈도, 백분율, 평균을 산출하였다.

연구 결과

1. 일반적 사항

총 189편의 연구 중 103편(53.5%)이 2000-2009년에 이루어졌으며, 2000년대 이후 도구 개발 연구가 급격히 증가하였다. 개발된 도구는 건강한 대상자에게 적용하는 도구가 81개(42.9%)로 가장 많았고, 환자 대상 50편(26.5%), 간호제공자 대상 45편(23.8%) 순이었다. 측정하고자 하는 개념으로는 정신사회적 개념이 50편(26.5%)으로 가장 많았으며, 건강지식 및 행위 31편(16.4%), 인지된 건강과 간호과정 및 간호수행이 각각 29편(15.3%)으로 뒤를 이었다. 171편의 연구가 도구 개발을 목적으로 하였으며 18편의 연구는 번역 도구의 표준화를 목적으로 수행된 연구였다(Table 1).

2. 타당도 유형

총 189편의 도구개발 및 타당화 연구에서 내용타당도, 구성타당도 및 준거타당도를 모두 보고한 연구는 28편(14.8%)이었으며, 내용타당도와 구성타당도를 보고한 연구는 78편(41.3%), 내용타당도와 준거타당도를 보고한 연구는 4편(2.1%), 구성타당도와 준거타당도를 보고한 연구는 5편(2.6%)이었으며, 나머지 연구에서는 한 유형의 타당도 근거를 보고하고 있는 것으로 분석되었다. 또한, 타당도의 유형에

대한 분류없이 그냥 타당도라는 포괄적인 용어를 사용한 연구가 6편(3.2%), 기타 타당도만을 보고한 연구가 6편(3.2%)이었다(Figure 2).

3. 내용타당도

대상논문 중 내용타당도에 대한 근거를 제시한 연구는 141편(74.6%)이었으며, 140편의 연구 중 가장 많이 사용한 분석 방법은 CVI (Content Validity Index)로 125편(89.3%)이었다. CVI로 분석한

Table 1. Characteristics of Studies (N=189)

Variables	Categories	n (%)
Year of publication	1970-1979	1 (0.5)
	1980-1989	3 (1.6)
	1990-1999	24 (12.7)
	2000-2009	103 (54.5)
	2010-2013	58 (30.7)
Participants in studies using the instruments	Healthy people	81 (42.9)
	Nursing providers	45 (23.8)
	Patients	50 (26.5)
	Family	13 (6.9)
Nursing concept	Functional health	13 (6.9)
	Physical health	11 (5.8)
	Psycho-social health	50 (26.5)
	Health related knowledge and behavior	31 (16.4)
	Perceived health	29 (15.3)
	Family health	14 (7.4)
	Nursing process and intervention	29 (15.3)
	Nursing organization	6 (3.2)
	Others	6 (3.2)
Purpose of studies	Instrument development	171 (90.5)
	Instrument standardization	18 (9.5)
Type of validity	Validity	6 (3.2)
	Content validity	141 (74.6)
	Construct validity	137 (72.5)
	Criterion validity	43 (22.8)
	Other validity	13 (6.9)

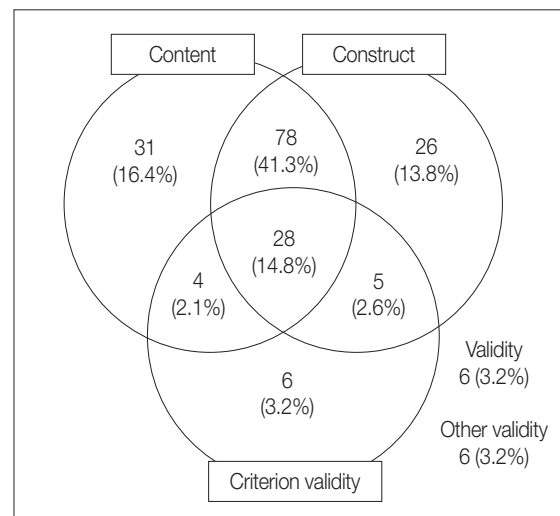


Figure 2. Proportion of reported validity.

연구에서 문항을 채택한 기준으로 가장 많이 제시한 수치는 80% 이상(58.1%)이었고, 연구에서 제시된 기준 CVI의 가장 낮은 수치는 50%였다(Table 2).

4. 구성타당도

구성타당도에 대한 근거를 제시한 논문은 137편(72.5%)이었으며, 검증을 위해 가장 많이 사용되고 있는 분석 방법은 탐색적 요인 분석으로 125편(91.2%)이었으며, 문항분석 53편(38.7%), 집단비교법 14편(10.2%), 수렴/판별 8편(5.8%) 순이었다. 구성타당도의 근거로 가장 많이 제시되고 있는 탐색적 요인 분석의 과정을 살펴보면 대부분 한 개 이상의 판단기준을 중복으로 제시하였으며, 요인의 개수를 판단하는 근거로 '고유값 1 이상'을 적용한 연구가 114편(83.2%)로 가장 많았으며 요인적재량 62편(47.4%), 스크리도표 25편(18.2%), 누적 분산설명량 22편(16.1%) 순이었다. 누적분산설명량을 기준으로 적용한 모든 연구가 60%를 기준치로 제시하였다(Table 2).

5. 준거타당도

총 189편의 연구 중 43편(22.7%)만이 준거타당도 검증을 실시했으며, 이들의 대부분(69.8%)은 연구를 통해 새롭게 개발된 도구와 다른 준거가 되는 도구와의 상관성이나 차이검증을 통해 동시타당도를 검증하였고, 11편(25.6%)의 연구에서는 동시타당도, 예측타당도, 수렴판별타당도의 구분 없이 준거타당도라는 포괄적 용어를 사용

하여 검증하였다. 준거타당도 검증방법은 79.4%에서 상관분석을 실시하였으며, 상관계수는 0.16-0.98까지 보고되었으며, 0.4 미만의 상관계수를 가지고 준거타당도를 검증한 연구 3편과 0.8 이상의 상관계수를 보고한 연구 4편이 있었다. 상관분석 외에 차이검정(17.5%), 회귀분석(1.7%), kappa (1.7%)도 사용되었다(Table 2).

6. 기타 타당도

도구의 기타 타당도를 검증하기 위해 연구자들은 민감도/특이도 7편(3.7%), 임상타당도 2편(1.1%), 집합타당도, 대조그룹타당도, 전문가타당도, Rasch모형을 통한 문항적합도는 각 1편씩 보고하였다. 임상타당도와 집합타당도는 다른 측정도구에 따른 집단별 차이분석이나 상관성분석을 실시하였으며, 민감도/특이도의 경우 선별(screening)을 목적으로 하는 도구개발 연구에서 주로 실시되었다(Table 3).

논 의

본 연구는 국내 간호학 분야의 도구개발 및 타당화 연구에서 보고되어온 타당도에 대한 동향을 파악하고 향후 도구개발 연구의 방향을 모색하고자 시도되었다. 분석 결과 간호학 분야의 국내학술지에 도구 개발 연구가 최초로 게재된 것은 1977년도였으며, 본격적으로 도구 개발 연구가 증가한 시기는 2000년대 이후임을 알 수 있다. 개발된 도구의 적용대상은 건강한 대상자가 42.9%로 가장 많았

Table 2. Type of Validity in Selected Studies

(N=189)

Type (n)	Variables	Categories	n (%)
Content validity (141)	Analytic methods	CVI	125 (89.3)
		Consensus of professionals	15 (10.7)
		Evaluation by discriminant group	1 (0.07)
	% of CVI (n=125)	Present	112 (89.6)
	Absent	13 (10.4)	
Construct validity (137)	Analytic methods	EFA	125 (91.2)
		CFA	4 (2.9)
		Item analysis	53 (38.7)
		Known-group validity	14 (10.2)
		Multitrait-multimethod	4 (2.9)
		Convergent/discriminant	8 (5.8)
		Evidence of deciding how many factors to extract in EFA	Eigen value > 1
		Scree plot	25 (18.2)
		Factor loading	65 (47.4)
		Cumulative variance	22 (16.1)
Criterion validity (43)	Subtype of criterion validity reported by researcher	Concurrent validity	30 (69.8)
		Predictive validity	1 (2.3)
		Convergent/discriminant validity	1 (2.3)
		Criterion validity	11 (25.6)

CVI=Content validity index; EFA=Exploratory factor analysis; CFA=Confirmatory factor analysis.

Table 3. Sensitivity and Specificity Studies

Authors (year)	Studied Phenomenon	Instruments	Samples	Speciality
Shin, Kim, & Lim (2007)	Developmental risks of 0-5 years old children	Korean-Child Developmental Review (K-CDR)	172 children aged 6-75 months old and their parents	Child
Shin, Kwon, & Lim (2005)	Developmental risks for high risk children	Korean version of Denver II Developmental Screening Test	113 children referred for developmental assessment in the department of rehabilitation of one university	Child
Choi et al. (2005)	Pressure ulcers	Pressure ulcers	73 inpatients aged 15 years and older in the neurosurgical department	Adult
Ju et al. (2009)	Developmental risks	Checklist for infant and child developmental screening	500 infants and children aged between 2 to 18 months	Child
So & Cho (2002)	Needs for home care	Home care need assessment tool	450 registered home care receivers	Community
Lee et al. (2011)	Frail status	Korea frail scale	632 seniors aged 65 years and older residing in the community	Gerontology
Kim et al. (2012)	Delirium	Korean Nursing Delirium Scale	361 older patients with delirium	Gerontology

다. 이는 2000년대 이후의 도구 개발 연구 동향을 분석한 Ribeiro 등 (2013)의 연구에서 임원환자를 대상으로 한 연구가 66.7% 였다는 결과와 비교해볼 때 국내에서 이루어진 도구 개발 연구가 초기에는 주로 질병특이적 개념보다는 일반적 개념을 측정하는 도구를 중심으로 이루어졌기 때문에 해석해 볼 수 있다. 측정하고자 하는 개념 중 정신사회적 개념이 26.5%로 가장 많았으며, 건강지식 및 행위 16.4%, 인지된 건강과 간호과정 및 간호수행이 각각 15.3% 였다는 결과가 이를 뒷받침한다.

측정하려는 도구가 측정하려는 개념을 얼마나 잘 반영하고 있는지에 대한 개념이 타당도이며, 학자들마다 이견은 존재하지만 내용타당도, 구성타당도, 준거타당도로 분류될 수 있다. 총 189편의 도구 개발 및 타당화 연구에서 가장 많이 언급된 타당도는 내용타당도로 74.6%를 차지하였으며 이는 Ribeiro 등(2013)의 연구 결과와 일치하는 결과이다. 한편, 내용타당도, 구성타당도 및 준거타당도를 모두 보고한 연구는 14.8%였으며, 두 가지 이상의 타당도 근거를 제시하고 있는 연구는 내용타당도와 구성타당도를 보고한 연구는 46.0%였다. 즉, 39.2%의 연구에서는 한 유형의 타당도 근거만을 보고하고 있는 것으로 나타났다. 이는 하나의 접근만으로 충분히 지지되는 강력한 타당도 근거는 없다고 봐야하며(AERA et al., 1999; Seong, 2002), 타당도가 수집된 모든 증거들이 도구의 점수에 대한 의도된 해석을 얼마나 지지하는지를 나타내는 통합된 개념으로 보아야 한다는 관점에서는 정말로 타당한 도구인가에 대한 비판을 피하기 어렵다. 즉, 도구타당화 연구에서는 해석적 논거의 추론과 가정에 부합하는 확보가능한 모든 증거를 수집함으로써 타당도의 논거가 명확해질 수 있도록 해야한다(Kane, 2000). 따라서, 일부 타당도의 근거를 통해 타당도가 '있다/없다'를 논하기보다 타당화의 다양하고 많은 논리적 분석을 제시함으로써 도구 사용에 대한 합리적 근거를 제공해야 한다.

한편, 본 연구의 분석 결과, 타당도에 대한 구분없이 포괄적인 용어를 사용하여 타당도를 검정한 연구들이 확인되었다. 즉, 세부적인 분류없이 타당도(6편, 3.2%) 혹은 준거타당도(11편, 5.8%)라는 포괄적인 용어를 사용하여 검증한 연구는 전체연구의 약 10%에 달했다. 이러한 경향성은 연구자들이 새롭게 개발된 도구가 어떤 측면에서 타당한지를 독자들과 앞으로 새롭게 개발된 도구를 사용하려는 연구자 및 임상가들을 논리적으로 설득하는데 충분한 정보를 제공하지 못하고 있는 것으로 생각된다.

국내에서 이루어진 도구 개발 연구에서 타당도의 근거로 가장 많이 제시되고 있는 근거는 도구의 내용에 기초한 근거로, 개발된 도구의 항목이나 문항이 측정하려는 속성을 측정하는 것으로 적합한가에 대한 내용전문가의 주관적인 판단을 분석하고 있다. 내용타당도는 도구의 목적에 부합하는지에 대한 중요한 의미를 가지고 있으나 공통된 인식이 없는 특성의 경우 각기 다른 검증 결과가 나오거나 계량화되지 않아 정도를 표기할 수 없다는 단점이 있다. 그러나 약 90%의 연구에서 CVI를 이용하여 내용타당도의 근거를 제시하였으며 대부분 4점 척도의 응답에 3점 이상이라고 응답한 비율이 80% 이상인 경우 해당 문항을 채택하는 기준으로 제시하였으며 (58.1%), 연구에서 제시된 기준 CVI의 가장 낮은 수치는 50%였다. 이러한 기준은 CVI가 0.50 이상인 경우 내용타당도의 수준이 받아들여질 만하다는 Waltz, Strickland와 Lenz (2010)의 기준에 비추어볼 때 타당한 수준임을 알 수 있다. 그러나 일부 연구에서는 CVI 결과를 척도 점수의 평균을 구한 값으로 제시하여 평균 3.39 혹은 3.72 등의 숫자로 제시하기도 하였으나 이는 CVI의 개념에 대한 부주의한 이해에서 비롯된 결과로 해석할 수 있다. 한편, 번역된 도구의 표준화가 목적인 경우 내용타당도를 제시하지 않은 연구도 있었으나, 타당도의 개념이 검사 결과의 해석의 적합성을 대변하는 것이라면 문화적 차이 등에 의해 도구 사용 시 고려해야할 부분이 있는 경우

번역 도구라 하더라도 내용타당도의 근거를 제시하는 것이 바람직할 것으로 사료된다.

구성타당도 혹은 구인타당도는 내적 구조에 기초한 근거를 의미하며 조작적 정의에서 규명한 구인을 제대로 측정하였는가를 검증한다. 대체로 상관관계나 탐색적 요인 분석, 확인적 요인 분석을 통해 근거를 도출하는 방법이 사용되며, 본 연구 결과에서도 탐색적 요인 분석을 통해 조작적 정의의 타당성을 제시한 연구가 66.1%로 가장 많았다. 탐색적 요인 분석을 실시한 125편의 연구 중 요인의 개수를 판단하는 기준으로 '고유값 1 이상'을 적용한 연구가 60.3%로 가장 많았으며 요인적재량(34.4%)과 스크리도표(13.2%), 누적분산 설명량(11.5%) 등을 고려하여 요인을 추출하였다. 이는 요인 분석 시 요인의 개수를 판단할 때, 고유값과 요인적재량, 스크리도표에서의 elbow와 함께 도구의 내용전문가인 연구자의 판단이 종합적으로 작용하여 판단한다는 기준(Pett, Lackey, & Sullivan, 2003)과 일치하는 것으로 대부분 통용되는 기준을 채택하고 있음을 알 수 있다. 한편, 내적 구조에 기초한 근거 자료를 제시하는 방법은 요인 분석만이 유일한 방법은 아니며, 그룹간의 비교, 상관계수를 이용한 수렴타당도 및 변별타당도의 검토 등이 사용될 수 있다(Kang, 2013). 또한, 건강관련 지식이나 인지적 측정도구의 경우 요인 분석 결과가 연구자가 의도한 바와 일치하지 않는 경우가 있어 이에 대한 해석이나 활용이 난해해지는 경우가 종종 있다. 이에 대해 Kang은 측정변수가 잠재변수의 값을 결정하는 구조인 형성적 지표를 이루는 측정변수들은 내적일관성 신뢰도 혹은 요인 분석의 대상이 되지 않는다고 밝히고 있다. 이와 관련하여 교육학적 연구에서는 인지적 평가도구의 경우 연구자가 의도한 평가목적과 피험자의 반응이 일치하는가에 대한 분석을 통해 반응과정타당도 근거(evidence based on response process)를 제시하는 것도 하나의 방법으로 사용되고 있으며(Hopfenbeck & Maul, 2011), 예를 들어 비판적 사고 능력이나 문제해결능력에 대한 도구 개발의 경우 개발된 문항이 추론 능력을 평가하기 위한 문항이라면 피험자가 사고 과정에서 추론을 이용하여 문항을 풀었는지에 대한 근거자료를 제시한다면 이 또한 타당한 도구임을 증명하는 주요 근거가 될 것으로 사료된다.

준거타당도란 새롭게 개발된 도구와 외적준거 사이에 높은 상관관계를 나타내는지를 검증하는 타당도로서 새롭게 개발된 도구와 외적준거의 측정치를 같은 시기에 측정하여 검증하는 동시타당도와 외적준거의 측정치를 미래에 측정하여 상관성을 살펴보는 예측타당도로 구분될 수 있다. 본 연구의 분석 결과 1편의 예측타당도 연구를 제외한 대부분의 연구에서 상관관계 검정을 사용한 동시타당도를 검증했으며 이때 보고된 상관계수는 0.16에서 0.98까지 보고되었다. 준거타당도를 검증할 때 상관계수는 0.4에서 0.8까지로 권장되며 낮은 상관계수는 새롭게 개발된 도구가 측정하려던 외적준거

와 다른 현상을 측정하고 있을 가능성을 시사한다(Lee et al., 2009; Park, 2005). 본 연구의 분석 결과 상관관계를 검증한 연구들 중 약 20%에서 새롭게 개발된 도구와 외적준거의 측정치 사이에 낮은 상관성을 보고하고 있어 새롭게 개발된 도구의 준거타당도가 수립되기 위해서는 앞으로 더 많은 연구가 필요함이 제기되고 있었다.

본 연구에서 기타타당도를 보고한 연구들 중 선별(screening)을 목적으로 하는 도구 개발 연구들은 비록 소수이기는 하나 의미가 깊고, 상대적 기준참조 평가방법보다 절대적 절단점(cutoff score)을 제시하는 준거참조 평가방법의 사용을 통해 도구의 목적이 적절히 달성되었다고 평가된다. 본 연구에서 분석된 연구들의 영역을 살펴 보았을 때, 정상아동과 발달문제를 지닌 아동의 판별, 임상적 증상의 심각도 및 서비스 적합도 등을 판별하기 위해 개발된 도구들이었으며, 이는 환자나 가족이 임상에서 진단을 받기 전 지역사회에서 pre-screening을 위해 사용되거나 고위험군을 대상으로 한 연구에 사용될 수 있어 유용하다. 간호학과 임상적 연결은 분리되어 생각될 수 없으며, 이러한 맥락에서 이미 기준참조 평가방법을 거친 도구들 또한 준거참조 평가방법을 사용한 타당화 과정을 거친다면 지역사회 건강전문가들과 연구자들에게 유용하게 사용될 것이다.

마지막으로, 도구 개발 연구를 수행하는 연구자들이 수렴/판별타당도를 일관되지 않게 사용하고 있었다. 수렴/판별타당도에 대해 연구자들은 세 가지 유형으로 범주화하였는데, 첫째는 내용타당도, 구성타당도, 준거타당도와 같은 수준의 새로운 범주(11편)로 파악하기도 하였으며, 때로는 구성타당도(8편)나 준거타당도(31편)의 일부로 파악하였다. 결과적으로, 도구 개발 연구를 하는 대다수의 연구자들은 수렴/판별타당도를 준거타당도의 일부로 파악하는 것을 알 수 있었다. 역사적으로, Cronbach와 Meehl (1955)은 타당도를 내용타당도와 구성타당도, 동시타당도와 예측타당도를 포함하는 준거관련타당도의 세 가지로 범주화하였고, 이러한 분류는 Campbell과 Fiske (1959)가 수렴/판별타당도의 개념소개를 포함하는 다속성/다방법 접근법을 새롭게 소개하면서 타당도에 대한 논의를 활발하게 만들었다. 그 후 학자에 따라 동일한 현상을 측정하는 기준도구나 외적 준거와 새롭게 개발된 도구와의 관련성을 구성타당도나 준거관련타당도에 포함시켜 구분하였으며, Devellis (2011)는 도구 개발자의 의도(intent)에 따라 분류하는 것이 바람직하다는 의견을 피력하였다. 이러한 맥락에서 대다수의 연구자들이 준거타당도나 구성타당도를 입증하는 방법으로 수렴/판별타당도를 검증하는 것은 이해될 수 있으나 수렴/판별타당도가 내용타당도, 구성타당도, 준거타당도와 같은 수준에서 논의되기 위해서는 이에 대한 연구자들의 부가적인 설명이 필요할 것으로 생각된다.

한편, 아직까지 국내 간호학 연구에서는 언급되지 않은 결과타당도 즉, 검사 결과에 기초한 근거(evidence based on consequences of

testing)도 추후 연구에서 적용될 수 있을 것이다. 특히, 간호교육연구에서는 시뮬레이션을 활용한 교육의 성과평가가 가져오는 의도한 결과와 의도하지 않은 결과, 긍정적인 결과와 부정적인 결과에 대한 분석이 가능할 것이다.

결 론

본 연구는 국내의 간호학연구 분야에서 도구 개발 및 타당도 연구경향을 살펴보고 앞으로의 연구방향을 제안하기위해 실시되었다. 분석 결과, 도구 개발 연구자들은 도구의 타당도를 입증하기위해 75%에서 두 가지 이상의 타당도 검증을 하고 있었으며, 개별타당도로는 내용타당도와 구성타당도를 주로 보고하고 있어 다양한 증거를 포함하는 타당도연구 및 개발된 도구와 도구가 측정하는 개념을 설명하는 이론과의 관련성을 입증하는 구성타당도가 점차 중요하게 여겨지는 추세를 반영하고 있어 고무적인 듯 보인다. 그러나 세부적으로 살펴봤을 때 구성타당도를 입증하는 증거의 대부분이 요인 분석에 그치고 있어 이론적 관련성을 설명할 수 있는 다양한 증거들을 포함하는 도구의 타당도 연구를 제언한다.

REFERENCES

- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (1999). *The standards for educational and psychological testing*. Washington, DC: AERA Publications.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105. <http://dx.doi.org/10.1037/h0046016>
- Carnevale, T. (2011). An integrative review of adolescent depression screening instruments: Applicability for use by school nurses. *Journal of Child and Adolescent Psychiatric Nursing*, 24(1), 51-57. <http://dx.doi.org/10.1111/j.1744-6171.2010.00256.x>
- Choi, S. H., Lee, E. J., Park, M. J., Kim, H. S., Kim, M. H., & Choi, J. Y. (2012). *Nursing outcomes measurement indicators*. Gwangju: Chonnam National University Press.
- Choi, S. J., Bae, S. Y., Choi, J. Y., & Bang, H. J. (2006). Development and predictive validity of pressure ulcer predicting scale for patients with neurologic condition. *Clinical Nursing Research*, 11(1), 95-108.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281-302.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications, Inc.
- DeVellis, R. F. (2011). *Scale development: Theory and applications* (3rd ed.). Newbury Park, CA: Sage Publications, Inc.
- Foster, R. L., & Park, J. H. (2012). An integrative review of literature examining psychometric properties of instruments measuring anxiety or fear in hospitalized children. *Pain Management Nursing*, 13(2), 94-106. <http://dx.doi.org/10.1016/j.pmn.2011.06.006>
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York, NY: Macmillan Publishing Company.
- Hopfenbecka, T. N., & Maul, A. (2011). Examining evidence for the validity of PISA learning strategy scales based on student response processes. *International Journal of Testing*, 11(2), 95-121. <http://dx.doi.org/10.1080/15305058.2010.529977>
- Ju, H. O., Lee, N. Y., Park, I. S., Lee, S. O., & Kim, S. H. (2009). Development & validation of a checklist for infant and child developmental screening. *Journal of Korean Academy of Child Health Nursing*, 15(1), 34-41. <http://dx.doi.org/10.4094/jkachn.2009.15.1.34>
- Kane, M. T. (2000). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. <http://dx.doi.org/10.1111/j.1745-3984.2001.tb01130.x>
- Kang, H. (2013). A guide on the use of factor analysis in the assessment of construct validity. *Journal of Korean Academy of Nursing*, 43(5), 587-594. <http://dx.doi.org/10.4040/jkan.2013.43.5.587>
- Kim, K. N., Kim, C. H., Kim, K. I., Yoo, H. J., Park, S. Y., & Park, Y. H. (2012). Development and validation of the Korean nursing delirium scale. *Journal of Korean Academy of Nursing*, 42(3), 414-423. <http://dx.doi.org/10.4040/jkan.2012.42.3.414>
- Lee, E. O., Im, N. Y., Park, H. A., Lee, I. S., Kim, J. I., Bae, J., et al. (2009). *Nursing research and statistical analysis* (4th ed.). Paju: Soomoonsa.
- Lee, I., Park, Y. I., Park, E., Lee, S. H., & Jeong, I. S. (2011). Validation of instruments to classify the frailty of the elderly in community. *Journal of Korean Academy of Community Health Nursing*, 22(3), 302-314.
- Lee, W. (2008). Unified view on validity. *Journal of Educational Evaluation*, 21(4), 67-79.
- Park, H. A. (2005). Problem and issues in developing measurement scales in nursing. *Journal of Nursing Query*, 14(1), 46-72.
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage Publications, Inc.
- Ribeiro, M. A. S., Vedovato, T. G., de Moraes Lopes, M. H. B., Monteiro, M. I., & de Brito Guirardello, E. (2013). Validation studies in nursing: Integrative review. *Rev Rene*, 14(1), 218-228.
- Seong, T. J. (2002). *Validity and reliability*. Seoul: Hakjisa Publisher.
- Shin, H., Kim, J., & Lim, S. (2007). A study to validate the Korean child development review. *Journal of Korean Academy of Child Health Nursing*, 13(2), 222-229.
- Shin, H., Kwon, B., & Lim, S. (2005). Validity of Korean version of denver II in screening children with developmental risk. *Korean Journal of Child Health Nursing*, 11(3), 316-321.
- So, A. Y., & Cho, B. H. (2002). Development of a home care need assessment tool: Focused on home care nursing diagnoses based on self care. *Journal of Korean Academy of Community Health Nursing*, 13(3), 433-443.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). *Measurement in nursing and health research* (4th ed.). New York, NY: Springer Publishing Company.