

외국 검사도구의 번역 및 타당화 과정에 대한 비판적 검토

—국어과 정의적 영역의 검사도구를 중심으로

장봉기 Oakland University

- I. 서론
- II. 동기 관련 검사지 번역 및 타당화 절차
- III. 연구방법
- IV. 연구 결과
- V. 논의 및 제언

I. 서론

학생들이 읽은 글의 내용을 이해하고 국어 시험에서 좋은 성적을 얻는 것은 국어과 학습의 중요한 목표이다. 이러한 인지적 목표 못지않게 중요한 것이 국어 공부를 좋아하고 국어 공부에 대한 목적을 스스로 설정하여 지속적으로 공부해 나가고자 하는 동기를 부여하는 정의적 영역에 대한 학습이다. 국어과 학습의 여러 동기관련 요소들을 학생들에게 효과적으로 가르치기 위해서는 학생들이 실제로 말하고 듣고 읽고 쓰는 과정에서 어떠한 정의적 특징을 보이는지 잘 평가하는 것이 중요하다. 평가에서 도출된 자료를 바탕으로 학생들의 국어과 학습동기를 이해하고 이를 국어수업에 반영할 수 있기 때문이다.

국어과 동기 관련 요소들을 평가하는 가장 일반적인 방법 중 하나는 이미 개발된 동기관련 검사지를 번역하여 타당도와 신뢰도를 검증한 뒤에 이용하는 방식이다(윤준채, 2008; Peña, 2007). 심리측정검사 개발에 드는 막대한 시간과 비용을 고려할 때, 이미 축적된 관련 이론을 바탕으로 개발된 도구를 번안하여 사용하는 것은 비용을 절감하고 개발 기간을 단축시키는 등

의 뚜렷한 장점이 있다. 하지만, 이러한 검사도구의 번역 및 타당화 과정이 심리측정이론에 근거한 정교한 절차를 따르지 않을 경우 발생하는 부작용도 만만치 않다. 특히, 리커트 척도(likert scale)로 구성되는 자기보고식의 검사지는 곧바로 수치화되어 도구 개발자의 의도나 목적과 관계없이 순위를 매기거나 중요한 교육적 의사결정에 직접적으로 활용되는 경우가 많기 때문에 설문문항의 제작과 타당도 검증에 각별히 주의해야 한다.

본 연구에서는 심리측정이론에서 제시하는 검사지 번역 및 타당화 절차를 바탕으로, 현재 국어교육 연구에서 수행되고 있는 검사지 번역 및 타당화의 현황과 생점을 살펴보려 한다. 이를 통해 국어과 동기 검사지 번역 및 개발의 정교한 절차를 확립하고 후속 연구와 교실 수업에 적용될 수 있는 여러 함의를 제시하고자 한다.

이를 위하여 본 연구에서는 다음과 같은 네 가지 연구문제를 설정하였다.

- 1) 심리측정이론에서 동기 관련 검사지의 번역 및 타당화를 위해 제시하는 정교한 절차는 무엇인가?
- 2) 국어과 동기 관련 연구에서 번역 및 타당도 검증이 이루어진 검사도구는 어떠한 것들이 있는가?
- 3) 국어과 동기 검사지 번역 및 타당화 연구에서 이러한 정교한 절차는 잘 적용되고 있는가?
- 4) 잘 적용되고 있지 않다면, 어떠한 세부절차가 간과되고 있으며, 이로 인한 문제점은 무엇인가?

II. 동기 관련 검사지 번역 및 타당화 절차

문식성 연구와 교수학습을 위해 그동안 수많은 동기관련 검사지들이 개발 및 사용되어 왔다. 새로운 심리측정 검사도구 개발 절차는 DeVel-

lis(2013), Henk, McKenna, and Conradi(2011), Netemeyer, Bearden, and Sharma(2003)에 의해 어느 정도 표준화된 절차가 제시되고 있다. 하지만, 동기 검사도구를 다른 문화권이나 언어권으로 번역하여 타당화하는 절차와 방법은 많은 연구자들이 공통된 기준 없이 상이한 방법을 사용하고 있는 실정이다. 이러한 문제의식을 바탕으로 Brislin, Lonner, and Thorndike(1973)는 심리검사도구를 다른 언어로 번역할 때 역번역과 이중언어 화자에 의한 검증, 공동번역, 사전검사 등의 절차를 반드시 포함시킬 것을 제안하였다. 같은 맥락에서 Maneesriwongul and Dixon(2004)은 심리측정 도구를 번역하여 사용한 47개의 연구를 검토한 뒤에 검사도구의 번역절차에 대한 표준적인 절차가 확립되지 못하고 있음을 지적하였다.

이에 본 연구에서는 문식성과 심리측정분야를 비롯하여 보건 및 간호학 등 다양한 학문 분야에서 다양적으로 연구되고 있는 검사지 번역 및 타당화 절차를 종합하여 국어과 학습과 관련된 동기 연구자들이 참고할 만한 표준화된 절차를 제시하고자 한다. 이를 위해, 앞서 언급한 여러 심리측정 문헌(DeVellis, 2013; Netemeyer, Bearden, & Sharma, 2003)과 다문화간 검사도구 번역 및 타당화 절차에 대해 심도 있게 논의한 여러 연구들(Borsa, Damásio, and Bandeira, 2012; Brislin, Lonner, and Thorndike, 1973; Byrne & Campbell, 1999; International Test Commission, 2010; OECD, 2013; Peña, 2007; Sousa & Rojjanasrirat, 2010; Tran 2009; Wild *et al.*, 2005)을 종합적으로 검토하여 〈표 1〉과 같은 상위 3단계, 하위 8단계 모형을 고안하였다. 각 단계에 대한 세부적인 설명은 아래와 같다.

표 1. 동기관련 검사도구 번역 및 타당화를 위한 8단계 모형

상위 단계	하위 단계	세부사항
번역	1. 준비단계	<ul style="list-style-type: none"> · 도구 개발자에게 연락을 취하여 번역에 대한 허가 요청 · 구인(construct)에 대한 개념 검토 · 번역 절차에 있어 핵심 인물들 접촉 및 확보
	2. 검사도구 번역 및 검토	<ul style="list-style-type: none"> · 개인 혹은 두 사람 이상의 연구팀원이 검사도구를 목표 언어로 번역하여 비교 및 조정
	3. 역번역 및 검토	<ul style="list-style-type: none"> · 목표 언어로 번역된 도구를 다시 원 언어로 번역하여 적합성 검토
내용 타당도 검증	4-1. 포커스그룹 인터뷰	<ul style="list-style-type: none"> · 검사도구의 목표 표본에 속하거나 표본에 대한 깊은 이해를 가진 관련 참여자들이 평가하고자 하는 구인의 개념과 각 문항에 대하여 그룹으로 토의
	4-2. 인지적 인터뷰	<ul style="list-style-type: none"> · 검사도구의 목표 표본에 속하는 소수의 참여자들에게 도구의 문항에 대한 인지적 이해 정도를 인지적 인터뷰를 통해 평가
	5. 전문가 검토	<ul style="list-style-type: none"> · 1차적으로 수정된 검사지를 해당 분야의 전문가 2인 이상에게 검토를 요청
구인타당도 및 신뢰도 검증	6. 예비 연구	<ul style="list-style-type: none"> · 검사도구의 타당도와 신뢰도를 검증하고 문항에서 발생될 수 있는 잠재적인 오류를 탐색하기 위해 예비 연구(pilot study) 실시
	7. 본 자료 수집 및 통계적 검증	<ul style="list-style-type: none"> · 검사도구 대상의 특징을 잘 대변할 수 있는 표본을 표집하여 검사도구의 타당도와 신뢰도를 평가하는 데 사용할 자료 수집 · 신뢰도 평가 · 차원성(dimensionality) 평가 · 구인타당도 평가
	8. 문항 조정 및 최종본 완성	<ul style="list-style-type: none"> · 통계적 검증을 바탕으로 검사지 재수정 및 최종 마무리

1. 준비 단계

가장 첫 번째 단계인 준비단계에서는 먼저 번역을 할 검사도구의 개발자에게 연락을 취하여 번역에 대한 허가를 구해야 한다. 학술지를 발간하는 출판사마다 특허권(copyright)을 강화하고 있는 추세이기에 공식적인 허가를 얻으면 이후에 발생할 수 있는 특허권 분쟁의 소지를 미연에 방지할 수 있다. 이러한 과정을 거쳐 번역 과정에서 발생할 수 있는 여러 문제를 해결

하는 데 원저작자의 협조를 구할 수도 있다.

한편, 번역과정에서 발생할 수 있는 오류를 줄이기 위해 개인 연구자가 혼자서 번역하는 것보다 번역팀이나 위원회를 꾸려 공동번역 작업을 진행하는 것이 더 효과적이다(Tran, 2009). 또한 준비단계에서는 번역 절차에 있어 핵심적인 역할을 할 구성원들을 접촉해서 확보하는 것도 중요하다. 주연구자와 함께 초벌 번역을 수행할 영어에 능통한 다른 연구자, 역번역을 수행할 이중 언어 사용자, 문항의 내용타당도를 평가할 전문가 집단, 포커스그룹 인터뷰와 인지적 그룹 인터뷰 참가자 등이 이러한 핵심적 구성원에 해당한다.

마지막으로, 본격적인 번역 과정에 앞서 검사도구가 측정하고자 하는 구인(construct)에 대해 광범위하고 체계적인 개념적 검토가 이루어져야 한다. 이를 통해 번역 과정에서 발생할 수 있는 개념적인 모호함을 방지하고 한 언어권 및 문화권에서 측정된 동기 관련 변인들이 또 다른 언어권에서도 타당하고 신뢰롭게 평가될 수 있는 토대를 마련할 수 있다(Borsa, Damásio, and Bandeira, 2012; Wild *et al.*, 2005).

2. 검사지 번역 및 검토

번역을 위한 사전준비가 모두 끝났으면, 2인 이상의 연구자가 원 검사도구를 개별적으로 번역한다. 이때 번역을 담당하는 연구자는 설문문항을 문자 그대로 직역하기보다 앞서 수행한 체계적 개념적 검토를 바탕으로 측정하는 구인의 특성에 맞게 맥락을 고려하여 번역해야 한다. 일차적으로 완성된 2개 이상의 번역본을 서로 비교·대조하면 작업은 문항 번역에 있어서 작용할 수 있는 개인 연구자의 편견이나 상이한 해석을 줄일 수 있다. 만약 1차적으로 수행된 2개 이상의 번역본에서 대립되거나 상이한 점이 발견될 경우는, 원 도구에서 제시하는 구인의 개념적 정의를 기준으로 번역자들 사이의 협의를 거쳐 번역 상의 합의를 도출한다.

한 연구자가 단독으로 검사도구를 번역하면 해당 연구자 개인의 언어

적 특징이나 습관, 오역 등이 여과 없이 그대로 검사지 문항에 반영될 위험이 커진다(Wild *et al.*, 2005). 예를 들어, Conradi, Jang, Lawrence, Craft, & McKenna(2013)가 개발한 청소년 독서태도 검사지를 한국어로 번역하여 타당화한 Jang, Ryoo, & McKenna(2013)를 보자. 학습 목적의 디지털 읽기 태도를 평가하기 위한 “How do you feel about reading a book online for a class?”라는 문항에 대하여 첫 번째 번역자는 “수업 준비를 위해 컴퓨터나 인터넷으로 책을 읽는 것에 대해 어떻게 느끼나요?”라고 번역한 반면, 두 번째 번역자는 “학습 목적으로 온라인으로 책을 읽는 것에 대하여 어떻게 생각하나요?”라고 번역하였다. 만약 이 번역과정에서 두 번역자 중 한 사람만이 번역을 했을 경우, 개별 연구자의 편견과 상이한 해석이 나타날 가능성 이 크다.

3. 역번역 및 검토

검사지를 1차로 번역하여 검토한 뒤에는 목표 언어(target language)인 한글로 번역된 도구를 원 언어인 영어로 다시 역번역하여 적합성을 검토해야 한다. 이때 번역의 타당성과 객관성을 유지하기 위해, 역번역을 수행하는 구성원은 1차 번역을 한 연구자들과 달리야 한다. 초기 번역 때와 마찬가지로 역번역자는 문자 그대로를 직역하기보다 구인의 개념적 특징을 잘 고려하여 역번역을 수행해야 한다. 이러한 과정을 거쳐 Brislin(1970)이 재안한 것처럼 원 언어와 목표 언어의 등가성(equivalence)을 평가하고 1차 번역에서 발생했을 수 있는 오류나 주관적 해석을 줄이도록 한다.

4. 전문가 검토

목표표본에 속하는 검사대상으로부터 포커스그룹이나 인지적 인터뷰를 통해 구인과 관련된 여러 가지 실제적 정보를 수집할 수 있다면, 해당 분

야의 전문가로부터는 좀 더 학술적인 내용타당도 검증을 받을 수 있다. 보통 해당분야의 석사 이상 학위 소지자 2인 이상에게 검토를 받는 것이 일반적이고, 검토 의견에 대해서는 번역위원회에서 숙고하여 검사지 수정에 반영한다. 전문가가 검토할 세부사항에 대해서는 주연구자와 번역위원회가 검사지의 특성을 바탕으로 협의해서 결정한다. Tran(2009)은 <표 2>처럼 문항진술의 명료성과 문항의 적합성, 문항난이도, 실제 맥락과 문항의 관련성의 네 가지를 평가하도록 제안한다.

표 2. 전문가 검토지 견본(Tran, 2009: 36)

검토자 성명:		
평가	적합 (이유)	부적합 (이유)
문항1		
명료성		
적합성		
난이도		
관련성		
문항2		
명료성		
적합성		
난이도		
관련성		
⋮		
문항		
명료성		
적합성		
난이도		
관련성		

5-1. 포커스그룹 인터뷰

포커스그룹 인터뷰는 검사도구의 대상자들이나 관련 인물들을 대상으로 좀 더 자연스러운 상황에서 구인과 관련된 좀 더 상세하고 실제적인 정보를 수집하기 위한 절차이다(Fowler, 2014; Groves *et al.*, 2009). 새로운 검사도구 개발에서 초벌 설문문항 제작에 중요한 역할을 하는 포커스그룹은 검사도구 번역에 있어서도 문화적, 언어적 차이에서 발생하는 이질성을 최소화하고 필요한 경우 문항을 삭제하거나 추가하는 의사결정의 중요한 토대를 제공한다.

이러한 포커스그룹 인터뷰는 주로 4~5명 내외의 참여자들이 한 그룹을 이루어 연구자가 미리 준비한 구인 관련 주제에 대해 자연스러운 분위기에서 의견을 나누는 방식으로 진행된다. 참여자는 검사도구의 직접적인 대상에 해당하거나 아니면 구인에 대한 실제적인 정보를 제공할 수 있는 간접적인 대상도 포함될 수 있다(Morgan, 1998). 예를 들어, 초등학교 고학년을 대상으로 제작된 동기 검사를 번역할 경우, 실제 초등학교 5학년 학생 서너 명을 대상으로 포커스그룹 인터뷰를 시행하여 번역된 동기 검사문항이 실제 학생들이 학교 안과 밖에서 수행하는 문식활동들을 잘 담고 있는지를 평가할 수 있다. 검사지의 대상이 초등학교 저학년일 경우는, 직접적으로 저학년 학생들을 대상으로 하기보다는, 간접적으로나마 저학년 학생들의 문식활동에 대한 정보를 제공해 줄 수 있는 부모나 교사들을 대상으로 하는 것이 바람직하다.

5-2. 인지적 인터뷰

역번역 검토와 포커스그룹 인터뷰까지 끝난 검사도구를 이용하여 목표 표본에 해당하는 소수의 참여자를 대상으로 인지적 인터뷰(Willis, 2005)를 시행한다. 이는 실제 검사도구를 접할 대상자들이 번역된 문항의 용어들을

도구개발자들이 의도한 바대로 잘 이해하는지를 확인하기 위함이다. 또한, 검사도구가 시행될 실제 대상자들이 설문문항을 이해하는 과정을 직접 살펴봄으로써 실제 참여자들이 좀 더 효과적으로 이해할 수 있는 방향으로 문항을 수정할 수 있다는 장점도 있다. 인지적 인터뷰는 목표 표본의 특징을 두루 가진 일반적인 대상에게 실시되어야 하며, 참여자 수는 필요에 따라 2~8명 내외로 한다. 검사도구와 구인, 그리고 검사대상의 특성에 따라 앞서 언급한 포커스그룹 인터뷰와 인지적 인터뷰 중 좀 더 적합한 형태의 인터뷰 하나만을 선택적으로 수행할 수도 있고, 두 가지 형태의 인터뷰 모두를 수행할 수도 있다(Henk, McKenna, & Conradi, 2011).

문화 간 비교연구자들이 보고한 바에 따르면, 특정 구인을 평가하는 문항이 새로운 문화권에서 다른 심리적 의미로 해석되는 경우가 많다. 예를 들어 윤준채(2008)에서 밝히는 것처럼 수업시간에 일어나서 급우들에게 책을 읽는 행위는 미국 학생들에게는 자연스러운 상황이지만, 한국의 초등학생들에게는 부담스러운 맥락으로 해석될 수 있어, 읽기 태도를 측정하는 문항으로는 부적합할 가능성이 높다. 이렇게 문화적 배경에 따른 해석의 다양성은 비가시적이므로, 인지적 인터뷰를 통해 얻을 수 있는 문항 이해과정에 대한 정보는 검사도구의 적합성과 관련성을 향상시키는 데 중요한 역할을 한다.

6. 예비 연구(pilot study)

전문가 검토와 포커스그룹 및 인지적 인터뷰를 통해 수정된 검사도구 번역본을 이용하여 검사도구의 타당도와 신뢰도에 대해 사전분석을 실시하기 위하여 예비 연구를 실시한다. 이 예비 자료는 본 검사에 앞서 검사도구 문항의 오류를 탐색하고 본 설문과정에서 발생될 수 있는 여러 가지 잠재적 문제들을 미리 확인하는 데 중요한 역할을 한다. 예비 연구 단계에는 편의 표집을 통해 작거나 중간 정도 크기의 표본을 대상으로 자료를 수집하는데, 보통 문항당 5명 정도의 표본 크기이면 적당하다(Netemeyer, Bearden, &

Sharma, 2003).

7. 본 자료 수집 및 통계적 검증

예비 연구에서 발견된 문제점을 보완한 뒤에 검사도구 대상의 특징을 잘 대변할 수 있는 표본을 체계적인 표집방법을 통해 선정하여 본 연구를 위한 자료를 수집한다. 이 자료는 검사도구의 타당도와 신뢰도를 평가하는데 중요하게 사용되므로 표본 크기가 예비 연구에 비해 커야 한다. 적절한 표본 크기의 기준은 여러 통계적 절차마다 다르지만, 보통 문항수가 많을수록 더 많은 표본이 필요하고, 300명 이상이거나(Nunnally, 1978) 전체 문항수의 열 배 이상의 표본이 적합하다(DeVellis, 2013)는 것이 심리측정 이론가들의 공통된 의견이다.

수집된 예비 자료를 바탕으로 검사도구의 신뢰도와 타당도를 평가하기 위해 문항분석, 요인분석, 그리고 신뢰도 분석의 세 가지 주요 통계 분석을 실시한다. 먼저 문항분석(item analysis)은 참여자들의 개별문항에 응답한 통계적 특성을 검증하는 단계이다. 보통 문항난이도(item difficulty)나 문항별 평균점수를 참고하여 개별문항의 질을 판단하는데, 널리 알려진 SPSS 프로그램이나 미국 버지니아대학교 Patrick Meyer 교수에 의해 개발된 jMetrick 프로그램(<http://www.itemanalysis.com/>)을 이용하여 좀 더 손쉽게 수행할 수 있다.

구인타당도는 먼저 상호상관을 통해 실제 같은 구인을 측정하는 문항 사이의 상관이 높게 나타는지를 검증하고 그런 다음 문항-총점 간 상관을 비교하여 개별문항 점수와 총점 간에 관련이 있는지 살피는 방식으로 진행된다. 이에 더하여 요인구조의 적합성을 판명하고자 요인분석을 실시하는데, 연구자가 구인의 이론적 구조에 대한 확신이 있을 때는 확인적 요인분석(confirmatory factor analysis, CFA)을, 그렇지 않을 때는 탐색적 요인분석(exploratory factor analysis, EFA)를 수행한다.

신뢰도 분석은 내적합치도(internal consistency)를 평가하는 것이 가장 일반적인데, Cronbach α 를 산출하여 평가가 가능하다. 이때 개별문항 점수와 총점간의 상관을 검토하고 신뢰도가 현격히 떨어지는 문항을 제외했을 때 나타나는 검사도구 전체의 신뢰도 변화를 검토한다. 문항 수에 따라 그 기준이 다르기는 하지만 일반적으로 동기관련 검사지의 경우 0.60 이상이면 신뢰로운 문항으로 간주할 수 있다. 이에 더하여 검사-재검사 신뢰도 역시 추천되지만, 동기관련 검사도구의 경우 기억이나 동기 요인들이 재검사의 점수에 영향을 미칠 수 있기 때문에, 선택적으로 수행하도록 한다(DeVellis, 2013).

8. 문항 조정 및 최종본 완성

앞선 단계에서의 타당도와 신뢰도 평가 결과를 바탕으로 연구자는 검사도구의 문항을 최종 수정하고 번역된 검사도구를 완성한다. 연구자가 한 가지 유의해야 할 것은 검사도구의 길이인데, 보통 짧은 검사도구일수록 검사대상이 느끼는 인지적 부담이 적은 장점이 있고, 긴 검사도구일수록 상대적으로 신뢰도가 높게 나타나는 장점이 있다. 결국 검사도구의 간결성과 신뢰도를 모두 고려하여 최적의 검사도구를 결정하는 것은 구인의 이론적 및 실제적 특성을 고려한 연구자의 판단에 달려 있다(DeVellis, 2013; Henk, McKenna, & Conradi, 2011).

III. 연구방법

본 연구의 첫 번째 연구문제와 관련하여 앞의 장에서 심리측정도구 개발에 대한 광범위한 문헌검토를 통해 동기 관련 검사지의 번역 및 타당화를

위한 체계적 절차를 살펴보았다. 다음으로 나머지 세 연구문제에 답하기 위하여 우리나라 동기 관련 국어교육 연구물을 아래와 같은 절차에 의해 수집 및 분석하였다.

1. 자료 수집

동기관련 국어교육 연구물을 수집하기 위해 가장 광범위한 검색결과를 제공하는 학술연구정보서비스(<http://www.riss.kr/>) 데이터베이스를 사용하였다. 검색의 효율성을 높이기 위해 세 가지 검색어를 동시에 입력하였는데, 제1 검색어는 “읽기”, “독서”, “쓰기”, “작문”과 같은 학습대상 관련 어휘를 사용하였다. 제2 검색어는 “동기”, “효능감”, “태도”, “흥미”, “자아개념” 등의 동기관련 어휘를 사용하였고, 마지막으로 “검사 도구”를 추가하여 관련문헌을 수집하였다. 이와 더불어 출판 범위는 국내학술지에 발간된 논문으로 한정하였다. 검색어 조합에 따른 검색결과는 <표 3>에 제시한 바와 같다. 1차적으로 수집된 논문자료를 검토하여 국어교육 관련 논문이 아니거나, 이미 번역된 검사도구를 재사용한 연구일 경우 분석의 대상에서 제외하였다. 이재승(2006)과 박영민(2008)처럼 다수의 외국 검사도구를 조합하여 새로운 검사도구를 개발한 경우 역시 최종 분석대상에서 제외하였다.

표 3. 국어교육 동기관련 논문 검색 결과

검색어 1	검색어 2	검색 결과물
읽기	태도	27
독서	태도	13
쓰기	태도	10
작문	태도	2
읽기	동기	17
독서	동기	9
쓰기	동기	8

작문	동기	2
읽기	효능감	8
독서	효능감	7
쓰기	효능감	5
작문	효능감	2
읽기	자아개념	5
독서	자아개념	6
쓰기	자아개념	2
작문	자아개념	0
읽기	흥미	19
독서	흥미	9
쓰기	흥미	7
작문	흥미	0
총 검색결과물		158

2. 자료 분석

본 연구의 목적에 부합한다고 판단되는 연구물을 최종 선별하여 〈표 4〉와 같은 분석틀을 마련하였다. 먼저 대상 연구물을 주저자의 성명을 기준으로 가나다 순으로 세로로 배열하고 앞서 정리된 8단계의 번역 및 타당화 절차를 순서대로 가로로 배열하였다. 연구물이 각 단계를 포함한 경우는 ○, 포함하지 경우는 ×, 그리고 부분적으로만 수행된 경우는 △로 부호화하였다. 분석과정에서 발생할 수 있는 오류를 최소화하기 위하여 연구자가 분석한 전체 자료의 30%를 독서교육 박사학위를 소지한 다른 연구자에게 똑같은 절차로 분석하도록 했다. 이 결과로 얻어진 평가자간 간 신뢰도(interrater reliability)는 0.98이었으며, 불일치한 부분은 추후 협의 과정을 통해 최종 합의에 이르렀다.

IV. 연구 결과

1. 번역 및 타당화된 동기 관련 검사도구

본 연구의 두 번째 연구문제에 대해 답하기 위하여, 동기 관련 국어교육 연구물을 종합적으로 검토한 뒤에 어떠한 동기 검사도구들이 번역되어 활용되고 있는지를 분석하였다. 총 10가지 동기관련 검사도구가 번역되어 활용되고 있었으며, 특징적인 사항은 Wigfield & Guthrie(1997)가 개발한 읽기 동기 설문지(Motivation for Reading Questionnaire, MRQ)가 다른 네 명의 연구자에 의해 개별적으로 번역되어 사용되었다는 점이다. 반면에 미국의 동기관련 연구와 교실현장에서 널리 쓰이고 있는 또 다른 동기 검사지인 Motivation to Read Profile(MRP, Gambrell, Palmer, Codling, & Mazzoni, 1996)은 번역되거나 타당화된 적이 없는 것으로 드러났다.

동기 관련 구인별로 살펴보면, 전체 10개의 번안 검사도구 중 절반이 조금 안 되는 4개의 도구가 태도를 측정하고 있었다. 효능감 역시 4가지 다른 도구가 번역되었다. 동기를 측정하는 도구는 MRQ가 유일했으며, 쓰기불안을 측정하는 도구가 하나 있었다.

표 4. 번역 및 타당화된 동기관련 검사도구 목록

연구물	동기 검사도구	측정 구인
권민균(2002)	Motivation for Reading Questionnaire (MRQ; Wigfield & Guthrie, 1997)	읽기 동기
기세령(2013)	Survey of Online Reading Attitudes and Behaviors and Skills (SOLAB; Putman, 2013)	읽기 태도
류보라(2013)	Reader Self-Perception Scale 2 (RSPS2; Henk & Melink, 2012)	읽기 효능감
문병상(2010, 2011)	MRQ (Wigfield & Guthrie, 1997)	읽기 동기
박영민·최숙기(2009a)	Pajares & Valiante (1997)	쓰기 효능감

박영민·최숙기(2009b)	Ohio State Teacher Efficacy Scale (OSTES; Tschanen-Moran & Hoy, 2001)	교사 효능감
박혜숙·전명남(2007)	MRQ (Wigfield & Guthrie, 1997)	읽기 동기
양연숙(2004)	Elementary Reading Attitudes Survey (ERAS; McKenna & Kear, 1990),	읽기 태도
	Elementary Reading Attitudes Survey (EWAS; Kear, Coffman, McKenna, 2000)	쓰기 태도
윤준채(2007)	Elementary Reading Attitudes Survey (ERAS; McKenna & Kear, 1990)	읽기 태도
윤준채(2009)	Kear, Coffman, McKenna (2000)	쓰기 태도
윤초희(2006)	Tullock-Rhody & Alexander (1980)	읽기 태도
최숙기(2008, 2009)	Reader Self-Perception Scale (RSPS; Henk & Melnick 1995)	읽기 효능감
최숙기(2010)	MRQ-Revised (Watkins & Coffey, 2004)	읽기 동기
	Estes attitude scale (Dulin & Chester, 1974)	읽기 태도
최숙기(2011)	Writing Apprehension Test (WAT; Daly & Miller, 1975)	쓰기 불안

2. 국어교육 동기 연구들의 번역 및 타당화 절차 평가

본 연구의 세 번째 연구문제에 대답하기 위해 앞서 살펴본 각 검사도구들의 번역 및 타당화 절차를 평가하였다. 세부적인 평가 항목과 결과는 <표 5>에 제시한 바와 같다. 분석에 쓰인 각 코드(○, △, ×)에 각각 1, 0.5, 0의 점수를 부여하여, 해당연구에서 여덟 단계를 얼마나 포함하고 있는지의 점수를 전체 연구물 수인 14로 나누어 %로 계산하였다. 그 결과, <표 5>에서 드러난 바와 같이 전반적으로 국어교육 연구에서 동기관련 검사도구들이 체계적 절차를 바탕으로 번역 및 타당화되고 있지 못함을 확인할 수 있었다.

이 결과를 바탕으로 네 번째 연구문제에 답하기 위해 각 검사도구들이 번역 및 타당화된 절차를 좀 더 면밀하게 살펴보았다. 먼저, 앞서 제시한 8가지 절차 중 포커스그룹 인터뷰는 어느 연구에서도 수행되지 않은 것으로 드러났다. 인지적 인터뷰 역시 4%로 거의 모든 연구에서 활용되지 않았다. <표

5>에서 알 수 있듯이, 기세령(2013)만이 유일하게 검사대상에 속하는 학생들에게 문항의 이해 정도와 어려움을 직접 물어 문항수정에 반영하였다. 하지만, 어떠한 방식으로 질문을 하고 문항수정에 반영했는지에 대한 언급이 없어 실제 이루어진 인지적 인터뷰 과정은 판단하기 어렵다. 그다음으로 낮은 수치를 보인 항목들은 역번역(21%)과 예비연구(21%)였다. 대부분의 연구에서 역번역 절차 없이 1차적으로 번역된 검사도구를 그대로 사용하였고, 검사문항의 오류나 검사도구의 신뢰도와 타당도에 대한 기초 자료를 확보하기 위한 예비연구 없이 바로 주연구를 위한 자료수집이 이루어졌다.

해당 분야의 전문가 검토 역시 36%밖에 이루어지지 않았고, 준비과정

표 5. 국어교육 동기관련 연구들의 번역 및 타당화 절차 평가

절차	준비	번역	역번역	인지적 인터뷰	포커스 그룹 인터뷰	전문가 검토	예비 연구	통계적 검증	수정 및 마무리
권민균(2002)	△	○	×	×	×	○	○	○	○
기세령(2013)	△	○	○	△	×	○	○	○	△
류보라(2013)	△	○	○	×	×	○	×	△	○
문병상(2010, 2011)	△	○	×	×	×	×	×	△	△
박영민·최숙기(2009a)	△	○	×	×	×	△	×	○	○
박영민·최숙기(2009b)	△	○	×	×	×	×	○	○	○
박혜숙·전명남(2008)	△	○	×	×	×	×	×	○	○
양연숙(2004)	×	○	×	×	×	×	×	△	△
윤준채(2007)	○	○	○	×	×	○	×	○	△
윤준채(2009)	△	○	×	×	×	×	×	○	△
윤초희(2006)	△	○	×	×	×	×	×	△	△
최숙기(2008, 2009)	△	○	×	×	×	×	×	△	○
최숙기(2010)	△	○	×	×	×	×	×	△	△
최숙기(2011)	△	○	×	×	×	△	×	○	○
총 연구물: 14	50%	100%	21%	4%	0%	36%	21%	79%	75%

○=1, △=0.5, □=0

에서 윤준채(2008)를 제외한 모든 연구들이 원 도구개발자에게 번역에 대한 허가를 구하지 않았다. 또한 대부분의 연구에서 번역된 도구에 대한 통계적 검증이 이루어졌지만 체계적인 검증의 정도는 모두 달랐다. 특히 구인의 차원성(dimensionality)을 검증하지 않은 경우가 두드러졌는데, 단순히 신뢰도 계수만을 보고하는 것은 해당 검사가 신뢰롭고 타당한지 체계적으로 검증하는 데 충분하지 않다. 마지막으로, 번역이 완성된 최종 검사지를 제시하지 않은 연구가 절반에 이르렀는데, <표 5>에서 수정 및 마무리 항목에 △표시가 된 연구들이 모두 이에 해당한다.

V. 논의 및 제언

본 연구는 윤준채(2008: 241)에서 제안한 것처럼, “외국의 도구를 번안하여 사용할 경우 번안된 검사도구의 양호도를 확보할 수 있는 타당한 절차들이 마련되어야 한다.”는 필요성에서 시작되었다. 그리하여 심리측정이론과 관련문헌을 종합적으로 검토한 뒤에 공통적으로 적용될 수 있는 번역 및 타당화 절차를 마련하였고, 이를 바탕으로 번안된 국어교육 동기관련 검사도구들의 유형과 번역 및 타당화 절차에 대한 쟁점을 살펴보았다.

단일 도구로서는 Wigfield & Guthrie(1997)의 독서 동기 검사지(Motivation for Reading Questionnaire, MRQ)가 네 연구자(권민균, 2002; 문병상, 2010, 2011; 박혜숙·전명남, 2008; 최숙기, 2010)에 의해 개별적으로 번역되어 가장 광범위하게 이용되는 것으로 드러났다. 이는 2003년 이후에 영어로 출판된 독서의 정의적 요인에 대한 양적 연구를 종합적으로 검토한 Conradi, Jang, & McKenna(2014)의 분석 결과와 일맥상통한다. 이들에 따르면, 분석대상이 된 총 81개의 연구 중 가장 많은 29%의 연구에서 MRQ나 수정된 MRQ가 사용된 것으로 드러났다. 미국의 교실에서 활발하게 사용되

고 있는 초등용 독서동기프로파일 검사지(Motivation to Read Profile, MRP; Gambrell *et al.*, 1996)나 청소년용 독서동기프로파일 검사지(Adolescent Motivation to Read Profile, AMRP; Pitcher *et al.*, 2007)은 번역되거나 타당화된 적이 없는 것으로 드러났다. MRQ에 비해 MRP는 초·중등용이 구분되어 있고, 설문문항에 더해 학생들의 독서동기와 관련된 좀 더 세부적인 정보를 묻는 인터뷰 질문까지 포함되어 있어 실제 교실상황에서 이루어지는 동기관련 연구에 활용가능성이 높다. 특히 초등용 MRP는 디지털 문식활동과 같은 초등학생들의 새로운 문식활동을 평가하기 위해 최근에 개정판(Malloy, Marinak, Gambrell, & MAzzoni, 2014)이 출판된 바 있다.

국어교육 동기관련 검사도구의 번역 및 타당화 절차에서 가장 뚜렷하게 드러난 결과는 인지적 인터뷰와 포커스그룹 인터뷰 절차의 누락이었다. 특히 인지적 인터뷰는 검사도구의 언어가 검사 대상자들에게 개발자의 의도대로 정확하고 잘 이해될 수 있는지를 검토하는 중요한 수단이기 때문에 이 과정을 누락하면, 검사 대상자들이 문항의 의미를 잘못 해석하거나 이해하여 부정확한 자료가 수집될 위험성이 커진다(Wild *et al.*, 2005; Wills, 2005). 또한, 만약 검사대상자가 문항 진술을 잘 이해하지 못하여 문항 반응을 포기하면, 이는 고스란히 결측치(missing data)로 전환되는 문제점이 발생한다 (DeVellis, 2013). 예를 들어, 초등학교 저학년 학생이 “단수, 복수, 시제, 접두사, 접미사를 정확하게 사용할 수 있다.”라고 번역된 쓰기 효능감 관련 문항을 읽었다고 가정해 보자. 쓰기에 대한 특정 과제를 수행할 수 있는지를 묻고 있는 이 문항이 쓰기 효능감을 측정하기에 적합하다는 것은 명확해 보이지만, 과연 초등학생들이 이 진술문에 사용된 언어를 이해할 수 있을지에 대해서는 의문이 든다. 이 경우 이 문항을 이용하여 수집된 검사 자료의 타당도와 신뢰도가 심각하게 손상될 수밖에 없다. 초등학교 6학년 학생들을 대상으로 번안된 또 다른 검사도구의 문항을 보자. “모든 독서과제를 완수해 내는 것은 나에게 매우 중요하다.”라는 문항은 독서 과제의 가치를 평가하기 위한 문항으로 적합해 보인다. 하지만, 만약 서너 명의 초등학교 6학년 학생

들에게 이 문항을 읽히고 인지적 인터뷰를 수행했다면, 어려운 한자어의 사용을 피하고 좀 더 명료하고 이해하기 쉬운 번역이 가능했을 것이다.

원 도구와의 개념적, 언어적 동질성을 확보하는 중요한 절차인 역번역 또한 빈번하게 누락되는 것으로 드러났다. 역번역 절차를 생략하면 오역이 발생할 가능성이 증가하는데, 한 예를 들면, “Correctly use all parts of speech in a written composition.”이란 문항은 “글을 쓸 때 모든 품사를 올바르게 사용할 수 있다.” 정도로 번역되는 것이 바람직하다. 하지만 실제 번안 도구에서는 “글을 쓸 때, 말하기의 모든 부분을 정확하게 사용할 수 있다.”라고 번역되어 언어적 등가성이 결여되었다.

Peña(2007)는 검사도구를 번역할 때 네 가지 등가성이 확보되어야 한다고 주장하는데, 언어적(linguistic), 기능적(functional), 문화적(cultural), 측량적(metric) 등가성이 이에 해당한다. 번역 - 연번역 과정을 통해 가장 직접적으로 향상시킬 수 있는 것이 바로 언어적 등가성인데, 원 검사도구와 번역된 검사도구에 포함된 어휘의 언어적 의미가 서로 동일한지를 평가하기 때문이다. 하지만 번역 - 역번역 과정을 통해 언어적 등가성을 확보하는 것으로 검사도구의 타당도를 위협하는 요소들을 모두 제거할 수는 없다. 이에 더하여 번역된 검사도구의 어휘들이 실제로 같은 기능을 하고 있는지 확인 할 필요가 있는데, 이를 기능적 등가성이라 한다. 예를 들어, Conradi, Jang, Lawrence, Craft, & McKenna(2013)가 개발한 청소년 독서태도검사를 보면, “How do you feel about being on social websites like Facebook or Myspace in your free time?”이라는 문항이 있다. 이 문항에서 “Myspace”는 미국 청소년들에게는 익숙한 소셜네트워크 서비스이지만 한국의 청소년들에게는 전혀 생소한 매체일 수밖에 없다. 이때 언어적 등가성만을 고집한다면, 문자 그대로 직역해야겠지만, 기능적 등가성을 고려한다면 한국 청소년들에게 좀 더 친숙한 “싸이월드” 정도로 번역하는 것이 내용 타당도를 유지하는 좋은 방법이다. 이러한 기능적 등가성은 “How do you feel?”을 “어떻게 느끼나요?”라고 직역하기보다 “어떻게 생각하나요?”나 “얼마나 좋아

하나요?”라고 의역하고자 하는 의사결정에도 적용될 수 있다. 측량적 등가성은 문항의 난이도를 동일하게 유지하는 것을 의미하는데, 예를 들어 위의 문항에서 번역자가 “social websites”를 소셜네트워크 서비스로 번역한다면, 원래 문항과 비교하여 번역된 문항의 난이도가 올라가리라는 예상을 할 수 있다. 미국 학생들에게 “social websites”라는 말이 아주 친숙한 반면, 우리나라 중학생들에게 “소셜네트워크 서비스”라는 말은 어려울 수 있기 때문이다. 마지막으로 문화적 등가성은 한 문항이 상이한 문화적 맥락에서도 동일하게 해석되는 것을 의미한다. 예를 들어, “수업 시간에 소설책을 읽는 것에 대해 어떻게 생각하나요?”라는 질문은 우리나라 중학생들에게 무척 응답하기 어려울 수 있는데, 이는 국어 수업에서 소설 전문을 읽는 경우가 그리 많지 않기 때문이다. 그래서 문화적 등가성을 고려한다면, “수업 시간에 교과서에 실린 소설을 읽는 것에 대해 어떻게 생각하나요?” 정도의 번역이 적합하다.

원저작자에게 번역에 대한 동의 및 협조를 요청하는 절차 역시 저작권을 존중하고 더 나아가 번역 과정에서 자문과 협조를 구하는 디딤돌이 될 수 있다는 점에서 아주 중요하다. 그리고 주연구를 위한 대단위 자료수집 이전에 작은 표집을 대상으로 한 예비연구 역시 현재보다 더 확대될 필요가 있다. 예비연구 단계에서는 신뢰도에 대한 정보를 미리 얻을 수 있고 이전에 개발된 관련 검사도구와의 상관도 살펴볼 수 있기에 그 중요성이 적지 않다. 대단위자료 수집이 완료된 뒤에 문항 오류를 발견하거나 기존의 검사도구와 너무 높은 상관이 드러나면, 또 다시 표집절차와 대단위자료 수집절차를 거쳐야 하는 막대한 시간적, 비용적 손실이 발생한다(DeVellis, 2013).

마지막으로 모든 절차를 거쳐 번역이 완성된 검사도구는 해당 연구가 타당화 절차를 포함하지 않더라도 후속연구를 위해 전문을 부록으로 실어 주는 것이 바람직하다. 본 연구의 분석대상이 된 14개 연구 중에 절반인 7개의 논문에서 번역된 도구의 세부적인 문항을 찾아볼 수 없었다. 더불어 검사도구를 사용하는 연구자와 교육자들을 위해 번역된 동기관련 검사지의 점

수를 세부적인 구인에 대한 고려 없이 총점으로 순위를 매기거나 다른 중요한 의사결정 과정에 단순하게 적용하지 않아야 한다는 사회적 타당도(social validity)의 측면도 언급 및 강조될 필요가 있다(조병영, 2011).

본 연구에서는 국어교육 연구에 있어 동기관련 검사도구의 번역 및 타당화의 현 주소를 살펴고, 좀 더 체계적인 8단계 번역 및 타당화 모형을 제시하면서 후속 연구들의 견고함을 높이는 데 기여하고자 하였다. 본 연구 결과 해석에 있어 개별 연구에 대한 질적 평가보다는 동기 관련 검사도구 번역에 대한 전반적인 경향을 살펴보는 것이 본 연구의 목적임을 다시 한 번 유념해 주길 바란다. 아울러 한정적인 검색어 사용으로 인해 특정 동기 검사도구가 자료 분석에 배제되었을 가능성이고 연구자가 제안한 번역의 방향 이외에 특수적인 상황이나 대상에 따라 더 적합한 대안들이 존재할 수 있음을 밝힌다. 이러한 본 연구의 시사점과 한계점을 바탕으로 후속 연구에서는 동기 관련 검사도구 번역 및 타당화에 대한 좀 더 견고하고 체계적인 논의를 기대한다.

* 본 논문은 2014. 1. 31. 투고되었으며, 2014. 2. 7. 심사가 시작되어 2014. 2. 28. 심사가 종료되었음.

참고문헌

- 권민균(2002), 「초등학생의 읽기동기에 관한 연구: 읽기동기의 구성요인, 학년과 성차이를 중심으로」, 『아동학회지』 23, pp. 17-33.
- 기세령(2013), 「인터넷 독서 성향 검사의 도구 타당도 검증」, 『독서연구』 30, pp. 227-259.
- 류보라(2013), 「고등학생의 독자로서의 자기 인식 연구」, 『새국어교육』 97, pp. 37-58.
- 문병상(2010), 「초등학생의 읽기 동기, 읽기 활동 및 읽기 능력간의 구조적 관계」, 『교육연구』 23, pp. 133-152.
- _____ (2011), 「내·외재적 읽기 동기, 자기조절 읽기, 읽기 활동과 읽기 능력의 관계」, 『중등교육연구』 59, pp. 123-143.
- 박영민·최숙기(2009a), 「우리나라 학생들의 쓰기 효능감 발달 연구」, 『새국어교육』 82, pp. 95-125
- _____ (2009b), 「국어교사 효능감의 구성 요인 분석」, 『학습자중심교과교육연구』 9, pp. 1-32.
- 박혜숙·전명남(2007), 「초등학교 6학년 대상 독서동기 검사의 타당화 연구」, 『교육심리연구』 21, pp. 969-988.
- 양연숙(2004), 「전래동화를 활용한 창의적 독서활동이 아동의 문해효능감과 정서능력에 미치는 영향」, 『한국영유아보육학』 39, pp. 129-147.
- 윤준채(2008), 「한국 초등학생 읽기 태도 검사 도구의 타당도 검증」, 『한국초등국어교육』 37, pp. 223-245.
- _____ (2009), 「초등학생 필자의 쓰기 발달 연구」, 『작문연구』 8, pp. 277-297.
- 윤초희(2006), 「언어영재의 문식적 환경, 독서태도, 지각된 언어능력 및 언어표현선호도가 쓰기 수행에 미치는 영향」, 『아동학회지』 27, pp. 53-68.
- 조병영(2011), 「구인타당도와 결과타동도 이론이 독서 평가의 해석과 사용에 주는 함의」, 『독서연구』 25, pp. 385-414.
- 최숙기(2008), 「인문계 고등학생 읽기효능감 구성 연구」, 『청립어문교육』 38, pp. 273-308
- _____ (2009), 「중학생의 읽기 효능감 구성 연구」, 『국어교육학연구』 35, pp. 507-544.
- _____ (2010), 「중학생의 읽기 동기와 읽기 태도에 관한 상관성 연구」, 『독서연구』 23, pp. 345-380.
- Borsa, J. C., Damásio, B. F., & Bandeira, D. R.(2012), "Cross-cultural Adaptation and Validation of Psychological Instruments: Some Considerations," *Paidéia(Ribeirão Preto)* 22, pp. 423-432.
- Brislin, R. W.(1970), "Back-translation for Cross-cultural Research," *Journal of Cross-Cultural Psychology* 1(3), pp. 185-216.
- Brislin, R. W., Lonner, W. J., & Throndike, R. M.(1973), *Cross-cultural Research Methods*, New York: John Wiley & Sons.
- Conradi, K., Jang, B. G., & McKenna, M. C.(2014), "Motivation Terminology in Reading

- Research: A Conceptual Review," *Educational Psychology Review* 26, pp. 127-164.
- Conradi, K., Jang, B. G., Lawrence, C., Craft, A., & McKenna, M. C.(2013), "Measuring Adolescents' Attitudes toward Reading: A classroom survey," *Journal of Adolescent & Adult Literacy* 56, pp. 555-566.
- DeVellis, R. F.(2013), *Scale Development: Theory and Applications* 3rd ed., Thousand Oaks, CA: Sage Publications.
- Fowler, F. J.(2014), *Survey Research Methods* 5th ed., Thousand Oaks, CA: Sage.
- Gable, R. K.(1986), *Instrument Development in the Affective Domain*, Boston, MA: Kluwer-Nijhoff.
- Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. (1996), "Assessing Motivation to Read," *The Reading Teacher* 49, pp. 518-533.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R.(2009), *Survey Methodology* 2nd ed., New York, NY: Wiley.
- Henk, W. A., McKenna, M. C., & Conradi, K.(2011), "Developing affective instrumentation for use in literacy research," In N. K. Duke & M. H. Mallette(Eds.), *Literacy research methodologies* 2nd ed., pp. 242-269, New York, NY: Guilford Press.
- International Test Commission(2010), *International Test Commission guidelines for translating and adapting Tests*, Available from <http://www.intestcom.org>.
- Jang, B. G., Ryoo, J. H., & McKenna, M. C.(2013, December), "Multiple Dimensions of Korean Adolescents' Reading Attitudes and Achievement," Paper presented at the annual meeting of the Literacy Research Association, Dallas, TX.
- Malloy, J. A., Marinak, B. A., Gambrell, L. B., & Mazzoni, S. A.(2014), "Assessing Motivation to Read: The Motivation to Read Profile-Revised," *The Reading Teacher* 67, pp. 273-282.
- Maneesriwongkul, W. & Dixon, J. K.(2004), "Instrument Translation Process: A Methods Review," *Journal of Advanced Nursing* 48, pp. 175-186.
- Morgan, D. L.(1998), *Focus Groups as Qualitative Research* 2nd Ed., Thousand Oaks, CA: Sage.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S.(2003), *Scaling Procedures: Issues and Applications*, Thousand Oaks, CA: Sage Publications.
- Nunnally, J.(1978), *Psychometric Theory*, Second Ed., New York, NY: McGraw-Hill.
- OECD(2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing.
- Peña, E. D.(2007) "Lost in Translation: Methodological Considerations in Cross-cultural Research," *Child Development* 78(4), pp. 1255-1264.
- Schwab, D. P.(1980), "Construct Validity in Organization Behavior," In Staw, B. M. & Cummings, L. L.(Eds.), *Research in Organizational Behavior*, Vol. 2, pp. 3-43, greenwich, CT: jAI Press Inc.
- Slavec A., & Drnovsek M.(2012), "A Perspective on Scale Development in Entrepreneurship Research," *Economic and Business Review* 14, pp. 39-62.
- Sousa, V., & Rojjanasrirat, W.(2010), "Translation, Adaptation and Validation of Instruments or Scales for Use in Cross-cultural Health Care Research: A Clear and User Friendly Guideline," *Journal of Evaluation in Clinical Practice* 17(2), pp. 268-274.
- Tran, T. V.(2009), *Developing Cross-Cultural Measurement.*" Oxford, UK: University

- Press.
- Wigfield, A., & Guthrie, J. T.(1997), "Relations of Children's Motivation for Rading to the Amount and Breadth of Their Reading," *Journal of Educational Psychology* 89, pp. 420–432.
- Wild, D., *et al.*(2005), "Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes(PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation," *Value in Health* 8, pp. 94-104.
- Willis, G. B.(2005), *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, Thousand Oaks, CA: Sage.

외국 검사도구의 번역 및 타당화 과정에 대한 비판적 검토

—국어과 정의적 영역의 검사도구를 중심으로

장봉기

본 연구의 목적은 심리측정이론과 문화 간 비교연구에서 제시하는 검사지 번역 및 타당화 절차들을 종합적으로 검토하고 이를 바탕으로 현재 국어 교육 연구에서 수행되고 있는 동기 관련 검사지 번역 및 타당화의 현황과 쟁점을 살피는 것이다. 관련 자료를 종합적으로 검토하여 총 8단계로 요약된 번역 및 타당화 절차가 마련되었다. 이 단계는 기존의 연구에서 자주 간과되었던 번역 준비, 역번역, 인지적 인터뷰, 포커스그룹, 예비연구 등을 두루 포함하였다.

이 단계를 바탕으로 수집된 관련 연구를 분석한 결과, 현재까지 번역 및 타당화된 국어과 동기관련 검사도구는 총 10종이었고, 특히 MRQ가 가장 빈번하고 반복적으로 활용되고 있는 것으로 드러났다. 세부적 절차와 관련하여 인지적 인터뷰 및 포커스그룹 인터뷰와 역번역 절차가 가장 두드러지게 생략되는 것으로 분석되었고, 전문가 검토와 예비연구 또한 좀 더 적극적으로 시행될 필요성이 제기되었다. 최근 국어과 정의적 요인에 대한 연구가 활발하게 증가되고 있는 점을 감안할 때, 본 연구는 관련 연구의 동향과 쟁점을 분석하고 검사도구 번역 및 타당화를 위한 종합적인 절차를 제시함으로써 체계적이고 견고한 방법적 토대를 확립하는 데 기여하리라 기대된다.

핵심어 동기, 검사도구, 번역, 타당화

ABSTRACT

Translation and Validation of Motivation Instruments in Literacy Research: Trends and Issues

Jang, Bong-Gee

The purpose of this study is to describe the importance of applying multiple validation procedures to translating and using motivation instruments in literacy research. I reviewed highly recommended methodological approaches to translation and cross-cultural validation of research instruments and established an eight-step guideline. This guideline includes multiple validation procedures such as seeking conceptual definition, using forward/backward translation processes, adopting group translation methods, conducting cognitive interviews and focus groups, encouraging written feedback on the survey, and using feedback on items from experts.

This guideline was used to evaluate literacy motivation instruments translated from English into Korean language. Results showed that Motivation for Reading Questionnaire (MRQ) was translated most frequently by four researchers and was identified as the most widely used motivation measure in literacy research in Korea. Cognitive interviews, focus groups, and back translation were often not conducted and this lack of methodological rigor created some major issues in validity and reliability of those translated instruments. Considering the increasing numbers of motivation studies in literacy, this study contributes to understanding methodological sensitivity and complexity in translating and validating motivation instruments.

KEYWORDS Motivation, translation, validation, measurement