

과학학습의 정의적 영역에서 사전-사후 통합 검사 설계의 타당화 연구: 과학영재를 대상으로

임채성 · 박형민[†]
(서울교육대학교)

A Validation Study of Retrospective Pre-post Test in the Affective Domain in Science Learning: for Scientifically Gifted Elementary Students

Lim, Chae-Seong · Park, Hyoung-Min[†]
(Seoul National University of Education)

ABSTRACT

In this study, the reliability and validity of the retrospective pre-post test were analyzed in order to solve the problem of traditional pre-post test including response shift bias. Samples of the study were 162 elementary school students who are studying at the S university gifted education center in Seoul. Before completion of the field trip, we conducted pre test of science-related attitudes. After completion of the field trip, respondents were asked to compare their responses of pre and post science-related attitudes to quantitatively analyze the commonalities and differences of the two tests. To find out more characteristics, qualitative data such as daily records and interview were also gathered and analyzed. The major results of the study are as follows. First, for the paired *t*-test, there was no statistically significant difference between separate pre-test scores and retrospective pre-test. There was a very high correlation between the separate pre-test scores and the retrospective pre-test. Second, there were significant differences in all seven sub-factors of science-related attitudes between the retrospective pre-test and the post-test. Third, the separate pre-test scores showed a slightly higher tendency than the retrospective pre-test scores. This suggests that the response shift bias appears when it is performed the separate pre-test in affective domain. As a result of the interview, it was found that the evaluation standards of separate pre-test did not match with those of post-test. Forth, internal consistency reliability of the retrospective pre-test was higher than that of the separate pre-test. However, there were significant differences in six factors of science-related attitudes excluding the 'social implications of science' between the separate pre-test and the post-test. Based on these results, the retrospective pre-post test design provides simplicity and convenience to both respondents and investigators, as it is done with one test. The retrospective pre-post test design can be regarded as a valid design for the self-report measurement of affective domain on a single experimental group.

Key words: science learning, affective domain, retrospective pre-post test, validation study, scientifically gifted elementary students

I. 서 론

본 연구는 과학학습의 정의적 영역에서 사전-사후 통합 검사 설계(retrospective pre-post test design)의 타당화를 위한 연구이다. 기존 교육 연구 분야에서 개발된 프로그램이나 자료가 학생들에게 미치는 영향을 평가할 때 일반적으로 정의적·행동적·인지적 영역을 망라하고, 사전-사후 검사 설계가 널리 활용되고 있다(Colosi & Dunifon, 2006; Dimitrov & Rumrill, 2003). 이 설계는 실험처치 전후에 검사를 2회 실시하여 처치효과를 검증하기 위한 연구 설계로, 고전적인 연구(Campbell & Stanley, 1963; Cook & Campbell, 1979)로부터 점차 정교화되어 왔다. 그러나, 이러한 사전-사후 검사 설계는 몇 가지 제한점이 있다. 첫째, 사전 검사와 사후 검사 사이의 시간 간격으로 인한 연구 시간 소모의 문제가 있다(Marshak *et al.*, 1998). 둘째, 참여자가 프로그램 투입 전과 후에 2회 반복 응답하게 하는 방식은 응답자의 응답 집중도와 충실도 감소와 같은 문제를 야기할 수 있다. 셋째, 사전 검사에서 응답변환편향 현상(response shift bias)이 나타나는 문제가 있다. 이러한 현상은 프로그램 투입 전 사전 검사를 실시할 때, 학생들이 자신의 태도, 기능, 지식 상태 등을 과대평가하여 사전 검사와 사후 검사 사이의 평가 기준이 모호해지는 경향을 의미한다(Howard & Dailey, 1979). 사전 검사에서 학생들의 상태가 과대평가된다면 처치를 통해 조사 대상자들에게서 나타나는 변화들을 과소평가하거나 왜곡시킬 수 있다(Hill & Betz, 2005; Pratt *et al.*, 2000, Rohs & Langone, 1997).

기존의 사전, 사후 분리 검사 설계는 비교 집단의 유무나 집단 동질성 통제 유무 등에 따라 단일집단·이질집단·통제집단 사전-사후검사 설계 등으로 분류된다(Fraenkel *et al.*, 2014; Gall *et al.*, 2006). 교육 연구 분야에서는 내적 타당도가 가장 높은 통제집단 사전-사후 검사 설계가 가장 널리 활용되고 있으며, 동일한 교육과정으로 모든 수업을 진행해야 하는 영재 수업 등 비교집단을 설정하기가 현실적으로 어려운 상황에서는 단일집단 사전-사후검사 설계도 활용되고 있다. 단일집단 사전-사후 설계는 하나의 실험집단을 대상으로 프로그램 처치 전과 후에 사전 검사와 사후 검사를 실시한다. 연구 설계의 종류에 따라 사전 검사의 성격은 달라지는데, 단일집단 사전-사후 검사 설계에서 사전 검사는 사후검

사와의 비교를 위한 준거검사(criterion test)의 성격을 지니고 있다. 그러나, 검사를 두 번 실시함으로써 피험자가 검사 내용을 기억하거나, 특정한 방식으로 응답하는 요령을 터득하는 등의 사전 검사 경험이 사후검사에 영향을 줄 수 있기 때문에 내적 타당도가 취약하다(성태제와 시기자, 2014; Dimitrov & Rumrill, 2003). 이러한 사전-사후 분리 검사 설계의 제한점을 극복하면서 연구의 타당성과 효율성을 높일 수 있는 검사 설계가 필요하다.

본 연구에서는 사전 검사 → 프로그램 처치 → 사후 검사로 이루어지는 단일집단 사전-사후 분리 검사 설계를 프로그램 처치 → 사전-사후 통합 검사 형태로 변형하여 그 타당성과 실용성을 검증하였다. 사전-사후 통합 검사 설계가 기존 설계와 주요하게 다른 점은 사전 검사와 사후 검사를 통합하여 실험 처치 이후에 1회만 검사를 실시한다는 점이다(Davis, 2003). 이처럼 검사를 1회만 실시하기 때문에 단일집단 사전-사후 분리 검사 설계에서 사전 검사의 경험이 사후 검사에 영향을 미쳐 야기될 수 있는 내적 타당도의 취약성을 보완할 수 있고, 사전-사후 통합 설계를 활용하여 응답 변환 편향 현상을 최소화할 수 있다(Marino *et al.*, 2010; Pratt *et al.*, 2000). 또한, 많은 연구자들이 직면하는 사전-사후 분리 검사 설계의 시간 제약의 문제를 해결하고(Lamb, 2005), 연구대상자들이 프로그램으로 인한 본인의 변화를 일관성 있는 기준을 가지고 평정하게 할 수 있는 장점들이 있다(Rockwell & Kohn, 1989).

사전-사후 통합 검사는 응답자가 참여한 프로그램에 비추어 자신의 지식과 행동의 실제적인 변화를 스스로 비교하게 하여 프로그램의 영향을 알아보고자 할 때 효과적이라고 보고되었다(Rockwell & Kohn, 1989; Stevens & Lodl, 1999). 본 연구는 행동적·인지적 영역 이외에 정의적 영역의 자기보고식 측정에서 사전-사후 통합 검사 설계의 타당성과 실용성을 검증한다는 점에서 의의가 있다고 할 수 있다.

이 연구에서는 의사소통의 편리성을 위하여 기존의 단일집단 사전-사후 분리 검사 설계의 사전 검사를 분리 사전 검사(separate pre-test)로, 사전 상태에 대한 사후의 통합 검사를 회고적 사전-사후 검사(retrospective pre-post test) 혹은 사전-사후 통합 검사(integrated pre-post test)라고 정의하였다. 영재학생들의 정의적 영역에 대한 분리 사전 검사 결과와 통합 사전 검사 결과는 어떠한 관계에 있으며, 분리

사전 검사 점수와 통합 사전 검사 점수의 차이점수의 유형에 따라 응답변환편향 현상에 차이가 나타나지 탐색하여, 통합 사전-사후 검사 설계의 타당성과 실용성을 검증하였다.

II. 연구 방법 및 절차

1. 연구 대상

본 연구는 서울시에 소재한 S교육대학교 과학영재교육원 10개 반(162명) 학생들을 대상으로 실시하였다. 사전-사후 통합 검사 설계는 주로 단일 실험 집단을 대상으로 한 연구에서 활용할 수 있고, 조사 대상자가 어떤 영역이나 특성에 대한 자신의 상태 변화나 차이를 반성적으로 파악하는 자기 고찰 능력이 어느 정도 갖춰져 있는 것을 전제로 하는 설계이기 때문에, 이에 적합하다고 판단되는 과학영재교육원에 재학 중인 학생들을 연구대상으로 선정하였다. 또한 본 연구자들이 해당 과학영재교육원의 반 지도교수 및 현장학습 인솔교사이기 때문에, 학생들의 특징이나 그 변화를 직접 관찰하기 용이하고, 심층적인 추가 면담 및 자료 수집이 가능하다고 판단하였다. 추가 면담은 응답변환편향 현상의 원인을 알아보기 위하여 실시하였으며, 분리 사전 검사 점수가 통합 사전 검사 점수보다 더 높은 유형의 학생들 중 차이점수가 평균보다 1SD 이상 높은 학생 2명을 대상으로 하였다.

2. 검사 도구

본 연구에서는 일반적인 과학 관련 태도 검사지로 널리 쓰이고 있는 TOSRA(Fraser, 1981)를 현장 학습의 성격과 내용에 맞게 수정하여 사용하였다. 과학교육을 전공하는 대학 교수와 대학원생들에게 의뢰하여 검사지의 타당도를 검증하였다. 이 검사 도구는 TOSRA 문항(70개) 중에서 하위 요인별로 타당도 평가 척도가 가장 높은 3개의 문항씩을 선정하여 5단계 리커트 척도형 21개 문항으로 구성하였다. 이 검사지는 과학의 사회적 의미, 과학자의 평범성, 과학 탐구에 대한 태도, 과학적 태도의 수용, 과학 수업의 즐거움, 과학에 대한 취미로서의 관심, 과학에 대한 직업으로서의 관심의 7개 하위 영역으로 구성되어 있다. 이 과학관련 태도 검사의 전체 문항에 대한 Cronbach α 값은 0.858이다.

3. 연구 설계

사전-사후 통합 검사 설계의 타당성을 검토하기 위하여 영재집단을 대상으로 현장학습을 실시하기 전에는 분리 사전 검사를 실시하고, 2박 3일간의 현장학습을 실시한 후에는 사전-사후 통합 검사를 실시하였다(Table 1). 이렇게 수집된 분리 사전 검사와 통합 사전 검사 데이터는 i-STATISTICS를 활용하여 대응표본 t -검정, 독립표본 t -검정, 일원 변량 분석, 상관 분석을 통해 정량적 분석을 실시하였다. 또한 정량적 검사지로는 명확히 드러나지 않는 특성들을 알아내기 위하여 참여 학생과의 인터뷰 자료를 분석하는 등 정성적 연구를 병행하였다(Creswell *et al.*, 2003).

III. 연구 결과 및 논의

1. 분리 사전 검사와 통합 사전 검사의 평균 비교와 상관분석 결과

전체적으로 과학 관련 태도에 대한 분리 사전 검사 점수가 통합 사전 검사 점수보다 높은 경향을 보였으나, 두 검사 점수의 차이는 통계적으로 유의하지 않았다 [$t(162, 162)=1.272(p>.05)$; Table 2]. 또한 상관 분석 결과, 두 검사 점수 사이의 상관계수는 0.92로 두 검사의 점수는 상관이 매우 높은 것으로 나타났다. 그러므로, 사전에 별도로 검사한 결과와 사후에 자신의 사전 상태를 회고하여 응답한 결과는 같다고 간주할 수 있다. 또한, 이 결과는 처치 후에 학생에게 그에 관련된 자신의 정의적 특성의 사전 상태

Table 1. Research design

연구절차		
O ₁	X	O ₂

O₁: 분리 사전 검사.
X: 과학 현장학습 프로그램.
O₂: 사전-사후 통합 검사.

Table 2. The result of independent sample t -test on science-related attitudes between separate pre-test and retrospective pre-test

집단	평균 (표준편차)	사례수	t
분리 사전 검사	4.55 (0.40)	162	1.272
통합 사전 검사	4.50 (0.38)	162	

와 사후 상태에 대해 동시에 응답하게 함으로써, 1회 검사로 사전 상태와 사후 상태를 모두 측정하는 방식이 타당하고 더 실용적임을 시사한다.

2. 통합 사전 검사와 통합 사후 검사, 분리 사전 검사와 통합 사후 검사 사이의 점수 차이 비교

통합 사전 검사와 통합 사후 검사의 과학 관련 태도 점수의 차이(0.15점)가 분리 사전 검사와 통합 사후 검사의 과학 관련 태도 점수의 차이(0.10점)보다 0.05점이 높았다. 또한 하위 요인별로 분석해 보았을 때, 과학 관련 태도의 7개 모든 하위 요인에서 통합 검사의 사전 점수와 사후 점수는 통계적으로 유의한 차이가 있었다(Table 3). 반면, 분리 사전 검사와 통합 사후 검사 점수의 경우에는 ‘과학의 사회적 의미’ 요인을 제외한 6개 요인에서 통계적으로 유의한 차이가 있다. 또한, ‘과학자의 평범성’ 요인을 제외한 6개의 하위 요인에서 분리 사전 검사 점수가 통합 사전 검사 점수보다 더 높았다. 이러한 결과들은 분리 사전 검사에서 학생들의 상태가 과대 평가되어 처치를 통해 응답자들에게서 나타나는 변화들이 과소평가되거나 왜곡될 수 있다는 선행 연구들의 분석과 일치한다(Hill & Betz, 2005; Pratt et al., 2000, Rohs & Langone, 1997).

3. 분리 사전 검사와 통합 사전 검사 점수의 차이 유형별 분석

분리 사전 검사 점수와 통합 사전 검사 점수의 차이점수의 유형에 따라 학생들을 3가지로 범주화 함으로써 유용한 정보를 찾아낼 수 있었다(Table 4). 첫 번째는 분리 사전 검사 점수가 더 높은 유형이다 (86명, 53.1%). 이 유형 학생들의 분리 검사와 통합 검사 차이점수 평균은 0.16점이며, 이 유형에 속하는 학생들의 분리 검사와 통합 검사 점수는 통계적으로 유의한 차이가 있다 $[t(86, 86)=3.089(p<.01)]$. 두 번째는 두 검사 점수가 같은 유형이다(17명, 10.5%). 세 번째는 통합 사전 검사 점수가 더 높은 유형이다 (59명, 36.4%). 이 유형에 속하는 학생들의 차이점수 평균은 -0.08 점이며, 분리 검사와 통합 검사 점수는 통계적으로 유의한 차이가 없다 $[t(59, 59)=-1.080 (p>.05)]$. 이 결과를 통해 분리 사전 검사의 점수가 더 높은 학생들이 더 많다는 것을 알 수 있으며, 정의적 특성을 자기보고식으로 측정할 때 분리 사전 검사에서 나타나는 응답변환편향 현상을 보여준다. 이러한 결과는 프로그램 처치 효과에 대해 분리 사전 검사와 통합 사전 검사 사이의 점수를 비교 연구한 선행 연구의 결과와 유사하다(Schwartz et al., 2004; Fig. 1). Fig. 1에서 알 수 있는 바와 같이, 사전-사후 통합 검사에서 나타난 효과가 사전-사후 분리 검사에서 나타난 효과보다 크다.

응답변환편향 현상의 원인을 알아보기로 하자 분리 사전 검사 점수가 더 높은 유형의 학생들 중 차이점수가 평균보다 1SD 이상 높은 학생들을 대상으로 면담을 실시하였다. 다음은 이에 관한 면담 내용 중 대표적인 두 가지 사례이다.

Table 3. Comparison of mean scores and change scores across three test types

	분리 사전 검사 점수 (표준편차)	통합 사전 검사 점수 (표준편차)	통합 사후 검사 점수 (표준편차)	통합 사후 검사와 분리 사전 검사의 차이 점수	통합 사후 검사와 통합 사전 검사의 차이 점수
전체	4.55(.399)	4.50(.382)	4.65(.340)	0.10***	0.15***
과학의 사회적 의미	4.74(.355)	4.70(.425)	4.79(.380)	0.05	0.09***
과학자의 평범성	4.29(.667)	4.42(.595)	4.57(.604)	0.28***	0.15***
과학 탐구에 대한 태도	4.50(.674)	4.44(.636)	4.60(.582)	0.10*	0.16***
과학적 태도의 수용	4.49(.503)	4.42(.517)	4.65(.412)	0.16***	0.23***
과학 수업의 즐거움	4.68(.504)	4.65(.510)	4.77(.431)	0.11***	0.22***
취미로서의 관심	4.53(.563)	4.45(.608)	4.61(.563)	0.07*	0.16***
직업으로서의 관심	4.39(.656)	4.38(.726)	4.54(.616)	0.15***	0.16***

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table 4. Comparison of change scores between separate pre-test and retrospective pre-test by types

유형	차이점수 (평균)	사례수 (명)	비율 (%)
분리 사전 검사 점수 > 통합 사전 검사 점수	0.16	86	53.1
분리 사전 검사 점수 = 통합 사전 검사 점수	0.00	17	10.5
분리 사전 검사 점수 < 통합 사전 검사 점수	-0.08	59	36.4

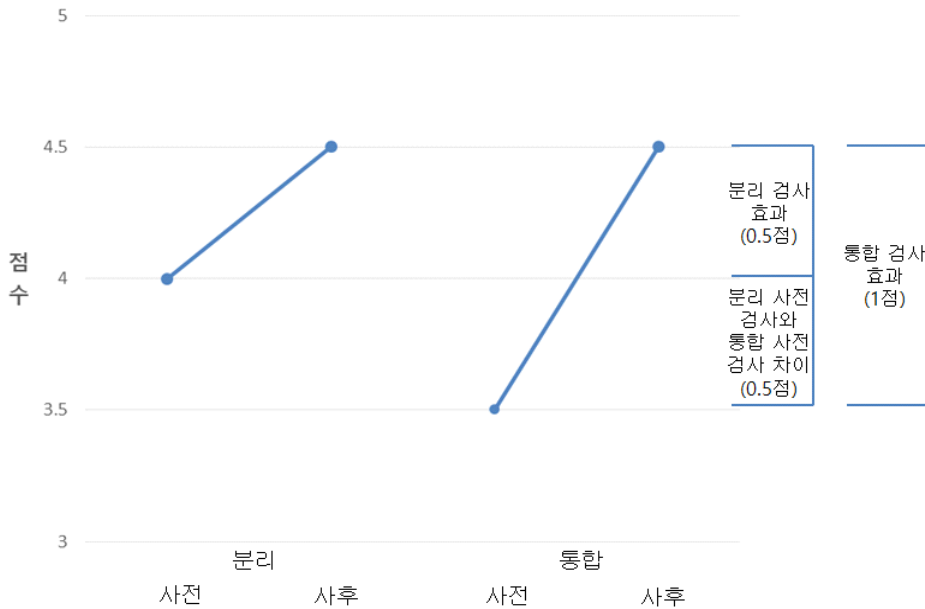


Fig. 1. Graphic representation of response shift and estimates of treatment effect using separate pre-test and retrospective pre-test (Schwartz *et al.*, 2004).

전*원: 현장학습 하기 전에는 내가 과학에 대해 어떻게 생각하는지 내 상태를 몰랐어요. 현장학습 활동을 하면서 내가 과학에 대해 어떻게 생각했었는지, 지금은 어떻게 생각하는지 알게 되었어요.

이 면담 사례의 밑줄 부분에서 알 수 있는 바와 같이, 응답자가 조사하려는 주제나 개념의 내용을 적절히 이해하지 못한 상태에서는 자신의 상태를 제대로 드러내지 못할 수 있다(Davis, 2003). 또한, 분리 사전 검사에서는 응답자가 프로그램에 참여하면서 본인의 사전 상태에 대한 응답이 틀렸음을 인지하였어도 이미 연구자에게 응답을 제출하였으므로 수정할 수 없다(Rockwell & Kohn, 1989). 이는 사전 검사와 사후 검사의 점수를 비교하기 위해서는 두 가지 검사를 평가하기 위한 공통적인 기준이 있어야 하는데(Cronbach & Furby, 1970), 분리 사전 검사에서는 응답자들이 사전 검사와 사후 검사 사이의 일관된 평가 기준을 갖기 어려운 점을 시사한다.

엄*빈: 현장학습 하기 전 검사를 할 때 점수를 실제보다 높게 썼어요. 내가 과학에 얼마나 관심이 있는지 선생님께 보여주기 위하여 전체적으로 높은 점수를 주었어요. 누구에게나 잘 보이고 싶은 생각이 있지 않나요? 내 점수가 높아야 남이 나를 신뢰해준다고 생각해요.

위 학생은 분리 사전 검사 때, 평가자의 평가를 의식하고 실제보다 더 높은 점수로 본인을 평가하였다. 이는 사회적 바람직성 편향(social desirability bias)이 나타났다고 해석할 수 있다. 사회적 바람직성 편향은 사람들이 프로그램의 목표에 맞추어 자신의 상태를 사회적으로 바람직한 방향으로 부풀리는 것을 의미한다(Anderson & Bourke, 2000; Colosi & Dunifon, 2006). 이 학생은 두 번째 검사인 통합 사전-사후 검사에서도 사회적 바람직성 편향을 일부 나타냈으나, 분리 사전 검사의 경우보다는 그 정도가 적었다.

이처럼 사전-사후 통합 검사 설계는 프로그램 처치 이전과 이후 사이의 평가 기준이 모호해지는 응답 변환 편향을 최소화할 수 있지만, 몇 가지 제한점이 있다. 위 면담 사례에서 알 수 있듯이, 타인에게 긍정적으로 보이게 하고 싶은 인간의 일반적 속성 때문에 사회적 바람직성 편향을 완전히 극복하기는 어렵다. 또한, 사전-사후 통합 검사는 학생들의 응답에 의존하기 때문에 일정한 기간이 소요되는 프로그램에 대한 자기 보고식 평가에서는 기억의 왜곡으로 인해 어떤 특성의 정확한 상태를 파악하기 어려울 수 있다(Schwartz *et al.*, 2004). 이러한 기억의 왜곡은 행동적 영역보다 정의적 영역의 평가에서 더 많이 나타난다(Schwartz & Rapkin, 2004). 따라서 정의적 영역에 대한 사전-사후 통합 검사는 학생들이 프로그램 처치 이전과 이후의 상태를 직접 비교하여 드러내기 때문에, 참여자가 인식하는 프로그램 효과를 평가할 때 더 적합하다. 그러나, 프로그램의 실제 효과를 파악하려면, 참여 관찰이나 산출물 평가 등 자료의 다각적 분석(triangulation)이 필요하다. 이에 대해서는 체계적인 후속 연구가 필요하다. 또한, 학생들의 성별, 전공 반, 선호하는 학습 유형이나 영역 등에 따라 응답변환편향에 차이가 있는지를 규명하는 연구도 필요하다.

4. 분리 사전 검사와 통합 사전 검사의 문항 내적 일관성 신뢰도 차이 분석

문항 내적 일관성 신뢰도란 검사를 구성하는 문항들이 측정하고자 하는 내용을 얼마나 일관성 있게 측정하였느냐 하는 정도이다(성태제, 2002; Henson, 2001). 연구 설계에 따라 문항 내적 일관성 신뢰도의 값은 달라질 수 있다(이학식과 김영, 1997; Churchill & Peter, 1984; Peterson, 1994). 그러므로 측정 자료의 신뢰도를 최대한 높일 수 있는 연구 설계의 탐색이 필요하다. 본 연구에서 비교하는 두 검사에 대한 문항 내적 일관성 신뢰도 검사 결과, 똑같은 과학 관련 태도 검사 문항이지만, 검사 시기가 다른

분리 사전 검사와 통합 사전 검사의 Cronbach α 값은 달랐다. 분리 사전 검사의 Cronbach α 값은 0.858, 통합 사전 검사의 Cronbach α 값은 0.867으로 통합 사전 검사의 신뢰도가 약간 더 높았다(Table 5). 또한 하위 요인별로 분석해 보았을 때, ‘과학 수업의 즐거움’ 요인을 제외한, 모든 요인에서 통합 사전 검사의 신뢰도가 약간 더 높았다. 이는 프로그램 처치 이후에 1회로 검사를 실시하는 사전-사후 통합 설계가 전통적인 분리 사전 검사 설계보다 신뢰도가 다소 높다는 점을 시사한다.

분리 사전 점수와 통합 사전 점수의 통계적 차이, 상관관계, 신뢰도 등을 종합적으로 고려할 때, 과학과 관련된 응답자의 정의적 특성에 대한 사전 상태를 프로그램 처치 이후에 통합으로 검사를 실시하는 것이 효율적이고 타당하다고 할 수 있다.

VI. 결론 및 제언

본 연구에서는 학생의 정의적 특성을 검사하는 사전-사후 분리 검사 설계의 응답변환편향 현상과 반복 측정으로 인한 주의집중력이나 응답성실도 저하 문제를 해결하고, 응답 효율을 높이기 위하여 사전-사후 통합 검사 설계의 신뢰도와 타당도를 분석하였다. 서울 소재 S교육대학교 영재교육원에 수학 중인 초등학교 학생 162명을 대상으로 과학 현장학습에 관한 영재과학캠프 참가 전에 사전 분리 과학 관련 태도 검사를 실시하고, 사후에는 응답자에게 사전과 사후를 비교하여 응답하게 하여 두 가지 검사 결과의 공통점과 차이점을 정량적으로 분석하였다. 또한, 정량적 검사로 드러나지 않는 특성들을 알아내기 위하여 관찰지, 면담일지를 분석하는 등 정성적 연구를 병행하였다. 본 연구를 통해 다음과 같은 결론을 도출하였다.

첫째, 과학 관련 태도 검사 결과, 분리 사전 검사 점수와 통합 사전 검사 점수 사이에는 통계적으로 유의한 차이가 없었고, 통합 사전 검사 점수와 분리

Table 5. Comparison of the Cronbach α coefficient between separate pre and retrospective pre-test scores

	과학의 사회적 의미	과학자의 평범성	과학 탐구에 대한 태도	과학적 태도의 수용	과학 수업의 즐거움	취미로서의 관심	직업으로서의 관심	전체
분리 사전 검사 (Cronbach α)	0.532	0.553	0.593	0.534	0.849	0.633	0.741	0.858
통합 사전 검사 (Cronbach α)	0.615	0.602	0.786	0.535	0.783	0.702	0.813	0.867

사전 검사 점수는 매우 높은 상관관계를 가지고 있었다.

둘째, 통합 사후 검사와 통합 사전 검사 사이에서는 과학 관련 태도의 7개 모든 하위 요인에서 유의한 차이가 나타났으나, 통합 사후 검사와 분리 사전 검사 사이에서는 ‘과학의 사회적 의미’ 요인을 제외한 6개 요인에서 통계적으로 유의한 차이가 나타났다.

셋째, 분리 사전 검사 점수는 통합 사전 검사 점수보다 약간 더 높은 경향을 보였다. 이는 분리 사전 검사의 경우, 과학의 정의적 영역에 대한 자기보고식 측정에서 응답변환편향 현상이 나타남을 시사한다. 면담 결과, 분리 사전 검사에서 학생들의 사전과 사후 검사에 대한 평가 기준이 일치하지 않았음을 알 수 있었다.

넷째, 통합 사전 검사의 응답이 분리 사전 검사의 응답보다 문항 내적 일관성 신뢰도가 더 높았다.

이러한 결과들을 종합하여 볼 때, 본 연구에서 활용한 사전-사후 통합 검사 설계는 사전-사후 분리 검사를 1회로 통합하여 실시함으로써 응답자와 조사자 모두에게 단순성과 편리성을 제공하고, 단일 실험 집단에 대한 프로그램의 정의적인 영향을 자기보고식으로 측정하기에 타당한 검사 설계라고 볼 수 있다.

이를 토대로 학교 현장에서의 실제적 적용 방안과 후속 연구 과제를 제안한다. 본 연구 결과는 정의적 영역을 자기보고식으로 측정하는 경우, 사전-사후 통합 검사 1회로 실시하는 것이 더 적절함을 시사하므로, 이를 토대로 일반 학생들을 대상으로 조사·분석할 필요가 있다. 또한, 과학탐구 영역과 과학 지식 영역에 대해서도 분리 사전 검사와 통합 사전 검사의 관계를 규명하는 연구가 필요하다. 본 연구에서는 단일 집단을 대상으로 사전-사후 통합 검사의 타당성을 탐색하였기 때문에, 두 집단 이상의 설계에서도 활용 가능한지에 대한 연구가 필요하다.

참고문헌

성태제 (2002). 타당도와 신뢰도. 서울: 학지사.
 성태제, 시기자 (2014). 연구방법론. 서울: 학지사.
 이학식, 김영 (1997). 연구디자인이 Cronbach's α 계수에 미치는 영향. 마케팅연구, 12(1), 209-221.
 한국교육심리학회 (2000). 교육심리학 용어사전. 서울: 학

지사.

Anderson, L. W. & Bourke, S. F. (2000). Assessing affective characteristics in the schools (2nd ed.). NJ: Erlbaum.
 Campbell, D. T. & Stanley, J. C. (1963). Experimental designs for research on teaching. *Handbook of Research on Teaching*, 171-246.
 Churchill Jr, G. A. & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. *Journal of Marketing Research*, 360-375.
 Colosi, L. & Dunifon, R. (2006). What's the difference: "Post the pre" & pre then post. Ithaca, NY: Cornell Cooperative Extension.
 Cook, T. D., Campbell, D. T. & Day, A. (1979). Quasi-experimentation: Design & analysis issues for field settings (Vol. 351). Boston: Houghton Mifflin.
 Creswell, J. W., Plano Clark, V. L., Gutmann, M. L. & Hanson, W. E. (2003). Advanced mixed methods research designs. *Handbook of Mixed Methods in Social and Behavioral Research*, 209-240.
 Cronbach, L. J. & Furby, L. (1970). How we should measure "change": Or should we?. *Psychological Bulletin*, 74(1), 68-80.
 Davis, G. A. (2003). Using a retrospective pre-post questionnaire to determine program impact. *Journal of Extension*, 41(4), 1-5.
 Dimitrov, D. M. & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work: Journal of Prevention, Assessment & Rehabilitation*, 20(2), 159-165.
 Fraenkel, J. R., Wallen, N. E. & Hyun, H. (9th ed. 2014). How to design and evaluate research in education. McGraw-Hill Higher Education.
 Fraser, B. J. (1981). TOSRA: Test of science-related attitudes: Handbook. Australian Council for Educational Research.
 Gall, M. D., Gall, J. P. & Borg, W. R. (8th ed. 2006). Educational research: An introduction. Pearson Education, Inc.
 Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development*, 34(3), 177-189.
 Hill, L. G. & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501-517.
 Howard, G. S. & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144-150.
 Lamb, T. (2005). The retrospective pretest: An imperfect

- but useful tool. *The Evaluation Exchange*, 11(2), 18-19.
- Marino, L., Lilienfeld, S. O., Malamud, R., Nobis, N. & Broglio, R. (2010). Do zoos and aquariums promote attitude change in visitors? A critical evaluation of the American zoo and aquarium study. *Society & Animals*, 18(2), 126-138.
- Marshak, H. H., De Silva, P. & Silberstein, J. (1998). Evaluation of a peer-taught nutrition education program for low-income parents. *Journal of Nutrition Education*, 30(5), 314-322.
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381-391.
- Pratt, C. C., McGuigan, W. M. & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pre-test methodology. *American Journal of Evaluation*, 21(3), 341-349.
- Rockwell, S. K. & Kohn, H. (1989). Post-then-pre evaluation: Measuring behavior change more accurately. *Journal of Extension*, 27, 19-21.
- Rohs, F. R. & Langone, C. A. (1997). Increased accuracy in measuring leadership impacts. *Journal of Leadership & Organizational Studies*, 4(1), 150-158.
- Schwartz, C. E. & Rapkin, B. D. (2004). Reconsidering the psychometrics of quality of life assessment in light of response shift and appraisal. *Health and Quality of Life Outcomes*, 2(1), 16.
- Schwartz, C. E., Sprangers, M. A., Carey, A. & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology & Health*, 19(1), 51-69.
- Stevens, G. L. & Lodl, K. A. (1999). Community coalitions: Identifying changes in coalition members as a result of training. *Journal of Extension*, 37(2), 1-9.