

평가 이해관계자들의 평가 의미 분석*

권 순 달**

수원대학교

본 연구는 ‘평가’에 대한 이해관계자들의 ‘평가’ 의미를 분석하는데 있다. 이를 위하여 평가의 개념에 대한 의미 분석(semantic differential methods)을 사용하였다. 이를 위하여 ‘평가’ 개념에 대한 3차원(활동차원, 능력차원, 평가차원)을 나타내는 형용사를 선정하고, 교사들에게 논리적 타당도를 검정하였다. 활동차원 12개 형용사, 능력차원 10개 형용사, 평가차원 4개 모두 26개 형용사를 사용하여 초중등학교 학생 및 대학생, 그리고 교사 등 604명을 대상으로 자료를 수집하여 탐색적 요인분석(EFA)을 하였다. 프로파일분석과 ANOVA를 실시하였으며, 구조관계를 파악하기 위하여 구조방정식모형으로 분석하였다. 연구결과 3차원 중에서 활동차원의 점수가 가장 높게 나타났으며, 초등학생을 제외한 중·고등학생, 교사 모두 능력차원에서 부정적으로 인식하고 있는 것으로 나타났다. 3개 차원의 요인계수는 모두 유의하게 나타났으며, 모형은 적합한 것으로 나타났다.

주제어 : 평가, 의미분석법, 구조방정식모형

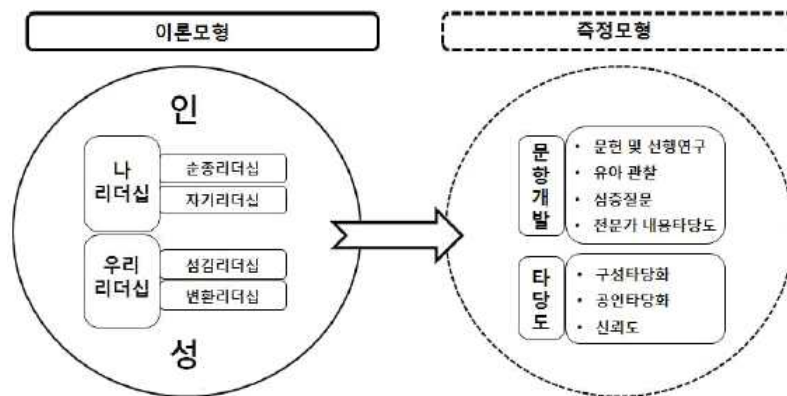
* 이 논문은 한국교육평가학회 2018 학술발표대회에 발표된 논문으로, 추가로 데이터를 수집하여 수정 작성됨.

** 교신저자 : 권순달, 수원대학교, sdkwon@suwon.ac.kr

I. 서 론

글을 쓰거나 논문을 작성할 때, 개념이나 용어를 정확하게 사용해야 하는 것은 아무리 강조해도 지나치지 않는다. 개념이나 단어의 의미가 명확해야 글의 내용이 독자에게 정확하게 전달되고, 혼란을 줄일 수 있다. 이러한 이유에서 석·박사 학위논문에서 용어의 정의를 활용하거나 각주를 사용하여 논문에서 중요하게 언급되는 용어를 정의하기도 한다.

최근 구조방정식모형으로 분석한 논문을 보면서, 학술연구 전문가들조차도 용어를 정확하게 사용하지 않고 있음을 발견하게 되었다. [그림 1]은 구조방정식모형에서 사용되는 ‘측정모형’의 용어를 사용하여 연구모형을 나타내고 있다(박영례, 2017, p.39).



(그림 1) 유아기 인성리더십 척도의 연구모형

[그림 1]에서 ‘이론모형’의 용어는 구조방정식에서만 사용되는 용어가 아니고 일반적으로 사용되고 있는 용어로 쉽게 이해되지만, 오른쪽의 그림을 나타내는 ‘측정모형’은 논문 내용으로 보면 ‘연구과정’, ‘연구절차’ 또는 ‘연구방법’으로 표현하는 것이 적절할 것이다. 그 이유는 ‘측정모형’의 용어는 구조방정식모형에서 측정변수와 이론변수의 구조 또는 관계를 나타내는 용어이기 때문이다.

이러한 이유에서 교육평가 전공자로서 교육평가 교직과목을 가르치면서 ‘평가’에 개념을 정확하게 사용하도록 노력하고 있으며, 학생들의 평가 결과 및 활용의 이해에 관심을 가지고 있었다. 그 이유는 ‘평가’의 개념에 대한 불분명한 이해는 ‘평가’에 대한 부정적인 인식과 함께 ‘평가’의 오·남용을 야기하는 것으로 밝혀졌기 때문이다(권순달, 2013). 그러므로 ‘평가’에 대한 부정확한 이해는 평가결과에 영향을 줄 수도 있다. 평가 이론가라고 할 수 있는 Alkin, Cousins, Patton 등은 ‘평가’에 대한 오·남용을 윤리적 관점에서 제기하였으며, 평

가 전문가에 있어서 중요한 문제라고 주장하였다(권순달, 2013). 그럼에도 불구하고 평가 전문가들조차 이러한 기본적인 내용에 별로 관심을 가지고 있지 않고 있다.

이러한 필요성에 따라 본 연구는 ‘평가’에 대한 이해관계자(stakeholder)들의 의미 차이를 연구하는 목적을 가지고 있으며, 연구 방법으로 전통적인 의미 분석적 방법(semantic differential method)에 따른 변화와 집단(초등학생, 중학생, 고등학생, 대학생, 교사)별 ‘평가’의 의미 차이를 파악하는데 있다. 학교급 별 집단을 비교하는 이유는 평가의 수행에서 초·중·고등학교가 다르기 때문이다. 초등학교는 중·고등학교처럼 평가가 빈번하게 실시되지도 않고, 평가 결과에 민감하지 않기 때문이다. 그리고 학교 현장에서 실제 평가를 담당하는 교사들도 중요한 평가 이해관계자라고 생각되었기 때문이다.

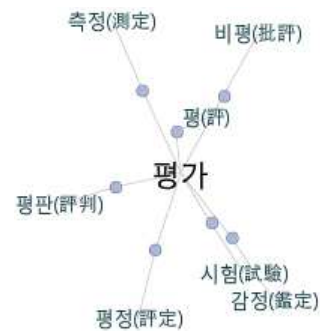
II. 이론적 배경

1. 평가 의미

‘평가’란 일련의 기준에 의해 대상의 장점, 가치 및 중요성에 대한 체계적 결정을 의미한다(Wikipedia, 2018). 평가는 무엇에 대한 질, 중요성, 양, 가치를 판단하거나 산정하는 것을 의미한다(Cambridge English Dictionary, 2018). ‘평가’란 대상에 대한 양, 수 또는 가치에 대한 판단을 내리는 것을 말한다(Oxford Dictionary, 2018). 이렇게 보면 ‘평가’란 사물에 대한 가치를 판단할 때 사용됨을 알 수 있다. ‘평가’의 이름으로 나오는 저널(EVALUATION)은 1995년에 SAGE에서 ‘평가’의 이론, 방법 및 실재를 향상시킬 목적으로 처음 발행되었다(SAGE, 2018). ‘평가’의 유사 용어로, 사정(assessment), 검사(examination, inspection, test, check), 추정(estimate, estimation), 판단(judgment) 또는 가치 판단(value judgment), 감정(appraisal, appraisement) 등이 있다. 사정(assessment)이 평가(evaluation)와 가장 비슷하게 사용되고 있지만, 이는 다음과 같이 구분할 수 있다. 평가(evaluation)는 기준과 증거에 근거하여 판단을 내리는 과정이며, 사정(assessment)은 학습에 관한 정보를 기술, 수집, 녹음, 채점 및 해석하는 과정이다(<http://tutorials.istudy.psu.edu/testing/testing2.html>, 2018)

그리고 검사란 사실이나 일의 상태 또는 물질의 구성 성분 따위를 조사하여 옳고 그름과 우수하고 우수하지 못함을 판단하는 일이다 그런데, 실제 학교 현장에서는 ‘평가’보다는 시험의 용어가 널리 사용되고 있다. 물론 시험과 ‘평가’는 전혀 다른 의미를 가지고 있다. 시험(examination)이란 인간의 능력이나 실력을 일정한 절차에 따라 검사하고 ‘평가’하는 일이다. 그러나 초·중고 학생들이나 일반인들은 ‘평가’를 시험의 동의어로 사용하고 있다. 그렇

지만 ‘평가’나 시험은 동의어가 아니다. 모든 시험을 ‘평가’라고 할 수 없으며, 모든 ‘평가’가 시험은 아니다. [그림 2]는 낱말 연구소에서 ‘평가’와 비슷하다고 생각되는 단어를 나타낸 것이다. 그렇지만 ‘평가’와 각 단어의 위치와 거리 관계가 무엇을 의미하는지 명확하지 않다. 한편, 네이버의 국어사전에서 ‘평가’와 비슷한 단어를 [그림 3]과 같이 나타내고 있다.



(그림 2) 낱말어휘정보처리연구소



(그림 3) 네이버 국어사전

[그림 2]와 [그림 3]에서 ‘평가’의 유사 용어의 관계를 시각적으로 나타내고 있으나, 거리 관계와 공간 위치 표시에 이론적 근거를 밝히고 있지 않고 있다. 그러므로 거리나 공간을 가지고 이들의 관계를 파악하는 것은 불가능하다. 예를 들면, 학교에서 많이 사용되는 ‘시험’의 단어도 [그림 2]와 [그림 3]에서 공간적 위치나 거리가 다를 수 있다. 일반적으로 교육 현장에서는 ‘평가’의 용어를 가장 널리 사용되고 있기 때문에 평가 의미를 분석하는 것이 타당할 것으로 판단된다.

2. 평가에 대한 연구

‘평가’에 관련된 선행연구가 다양하게 이루어져 왔다. 그러나 이러한 연구들은 ‘평가’ 그 자체의 개념이나 의미에 대한 연구보다는 ‘평가’와 관련된 인식이나 태도를 측정하는 연구가 대부분이었다. 이와 함께 ‘평가’ 불안(또는 시험 불안)에 대한 연구와 측정 도구도 제작되어 활용되고 있다.

‘평가’ 관련 선행 연구를 정리하면 크게 4 가지로 요약할 수 있다. 첫째, ‘평가○○’에 대한 연구가 있다. 이는 ‘평가’의 정의적 특성에 대한 연구라고 할 수 있다. 최근 이와 같은 연구로 고등학생의 ‘평가’인식 유형 분석(조영민, 2017)과 ‘평가’에 대한 태도 척도(SATA-Q) 연구(노현중, 2017), 초등학생용 ‘평가’에 대한 태도 척도(SATA-Q) 개발 및 타당화(노현중, 손원숙, 2018) 등이 이러한 유형의 연구라고 할 수 있다.

둘째, 자기평가, 수행평가, 형성평가, 진단평가 등 ‘○○평가’와 같은 평가유형에 대한 연구라고 할 수 있다. 학생 자기평가의 학습효과에 관한 메타분석(홍소영, 2018), EFL 학습자의 자기평가에 대한 인식 변화 분석: 사회문화적 관점에서(표경현·나경희, 2013) 등이 이런 유형의 연구이다.

셋째, ‘평가○○’을 독립변수로 하여 다른 종속변수에 영향을 미치는 연구가 있다. 평가인식 및 다른 여러 변수를 독립변수로 선정하여 다른 종속변수에 영향을 미치는 인과연구(이영만·하봉운, 2015), 이영만(2015)가 이런 유형의 연구라고 할 수 있다.

넷째, ‘평가○○’을 종속변수로 다른 변수의 영향력을 분석한 인과연구가 있다. 마을 만들기 사업의 참여경험이 평가인식에 미치는 영향(이재완, 2017), 상사의 배려행동, 상사-부하 간 교환관계가 조직시민행동 및 성과 평가인식에 미치는 영향(이종찬, 2009) 등이 그 예이다. 특히 평가○○을(를) 독립변수로 수행한 연구보다는 ‘평가○○’가 종속변인으로 수행된 연구와, ‘평가○○’ 그 자체에 대한 연구가 더 많은 것을 알 수 있다.

이와 함께 ‘평가’의 오용과 남용에 관한 연구를 통해 ‘평가’의 개념을 정확하게 이해할 필요가 있다는 연구도 수행되었다(권순달, 2013). 그러나 이 연구는 ‘평가’ 개념 분석이 아닌 오·남용 유형을 세 가지 유형으로 분석하는데 그쳤기 때문에 ‘평가’의 개념에 대한 분석이라고 할 수는 없다.

이렇게 보면 대부분의 연구들이 ‘평가○○’에 대한 연구와 평가유형, 평가 관련 인과연구이며, 평가의 개념 그 자체 즉 이미지에 대한 연구는 수행되지 않고 있음을 알 수 있다. 그러나 어떤 용어 또는 개념 즉, 이미지에 대한 연구는 다양한 분야-언어학, 심리학, 사회학, 예술 등-에서 이루어지고 있다(송영옥·최혁준, 2017; 이상미, 2016; 윤희성,

2016; 최종석, 2015; 조윤교, 2010). 따라서 본 연구는 ‘평가’의 개념을 의미 분석하는데 그 의의를 가지고 있다.

III. 연구 방법

1. 측정도구

특정한 개념이나 이미지에 대한 연구 방법으로 투영법, 심층면접법, 의미분석법 등이 사용되고 있다. 이 중에서 의미분석법(semantic differential)이 개념에 대한 의미를 연구하는 널리 사용되고 있다(송영욱·최혁준, 2017). 이 방법은 여러 가지 사물, 인간, 사상 등에 관한 개념이나 의미를 의미공간에서 측정하는 방법이다(황정규, 2002). Osgood은 50개의 상반된 의미를 갖는 형용사를 제안하였으며, 이러한 형용사들은 크게 평가차원(evaluative factor), 능력차원(potency factor), 활동차원(activity factor) 등 각각 독립된 3차원의 의미공간으로 구분되고 있다(성태제, 2016; 성태제, 2014; 권순달, 2011). 그런데 형용사를 선택할 때는 측정하려는 개념의 대표성 또는 측정 가능성을 고려해야 한다(송영욱·최혁준, 2017). 일반적으로 형용사들을 평가요인, 능력요인, 활동요인으로 구분하고 있으나, 이러한 요인은 연구의 목적과 개념에 따라 조정될 수도 있으며, 다른 요인을 첨부하거나 일부만 선택하여 사용할 수도 있다(Ding & Ng, 2008). 본 연구에서 사용된 측정도구는 1957년 Osgood에 의하여 제안한 3차원을 표현하는 형용사군에서 ‘남성적-여성적’ 등을 제외하고, ‘평가’의 의미를 표현하는데 적합하다고 생각되는 형용사-유용성-를 일부 포함시켜 <표 1>과 같이 분류하였다.

〈표 1〉 ‘평가’ 의미 분석에 사용되는 형용사

활동차원	무단-섬세한, 낯은-새로운, 수동적-능동적, 불성실한-성실한, 필요하지 않은-필요한, 위험한-안전한, 무거운-가벼운, 익숙하지 않은-익숙한, 도움되지 않는-도움이 되는, 부담이 되는-부담 안 되는
능력차원	복잡한-단순한, 공정하지 않은-공정한, 불안한-편안한, 까다로운-까다롭지 않은, 실용적이 아닌-실용적인, 부정확한-정확한, 공격적-방어적, 믿을 수 없는-믿을 수 있는, 어려운-쉬운, 흥미롭지 않은-흥미로운
평가차원	더러운-깨끗한, 나쁜-좋은, 옳지 않은-옳은, 가치 없는-가치 있는, 재미없는-재미 있는, 불완전한-완전한, 고맙지 않은-고마운, 쓸모없는-쓸모있는, 중요하지 않은-중요한, 즐겁지 않은-즐거운

위의 <표 1>에서 부정적 의미의 형용사들을 왼쪽, 긍정적 의미의 형용사들을 오른쪽으로 배치하여, 두 형용사 사이를 5점 척도로 응답하게 할 수 있도록 [그림 4]와 같이 질문지를 구성하였다.

복잡한					단순한
⋮	중략				⋮
즐겁지 않은					즐거운

(그림 4) '평가' 의미 분석 질문지 구성(예)

1) 측정도구 타당도

5명의 현직 교사들로부터 30개 형용사에 대한 논리적 타당도를 검증받아 '평가'의 개념을 표현하기에 적절하지 않은 '공격적-방어적', '무단-섬세한', '낯은-새로운', '수동적-능동적' 등 4개의 형용사를 제외하여 최종적으로 26개 형용사를 사용하여, 탐색적 요인 분석(EFA)을 실시하였다. 요인 수를 찾기 위하여 주성분과 함께 베리맥스(Varimax)방법으로 회전하였으며, Eigen Value 1이상으로 요인을 추출하였다. 그 결과 평가차원(evaluative factor)의 5개 형용사가 활동차원으로, 2개 형용사가 능력차원으로 묶였으며, 활동차원의 '위험한-안전한' 형용사가 평가차원으로 이동되어 <표 2>와 같이 3요인으로 분석되었다.

위의 <표 2>에서 요인1은 활동차원의 형용사 12개, 요인2는 능력차원의 형용사 10개, 요인3은 평가차원의 4개 형용사로 범주화하였다. 요인분석 결과 차원별 형용사들이 이동된 이유는 '평가'의 개념이 추상명사이기 때문이었다. 따라서 일부 형용사를 차원별로 명확하게 구분하는 것이 쉽지 않음을 알 수 있었다.

2) 측정도구 신뢰도

'평가' 의미 형용사들의 3 요인 즉 활동차원, 능력차원 그리고 평가차원에 대한 신뢰도 분석 결과는 <표 3>과 같다.

위의 <표 3>에 의하면 3 차원의 각 문항의 신뢰도는 매우 높은 것으로 나타났다.

2. 연구대상

본 연구의 대상은 '평가' 관련 이해관련자(stakeholder)라고 생각되는 경기도교육청과 서울특별시교육청 관내 초등학생, 중학생, 고등학생, 교사를 대상으로 2018년 4월 5일부터 5월 4일까지 1개월 동안 자료를 수집하였다.

〈표 2〉 탐색적 요인분석 결과

형용사	성분		
	1	2	3
복잡한 - 단순한	-.112	.439	.154
공정하지 않은 - 공정한	.660	.002	.002
불안한 - 편안한	.063	.676	.003
까다로운 - 까다롭지 않은	-.114	.669	.089
실용적이 아닌 - 실용적인	.625	.223	.144
부정확한 - 정확한	.755	.107	.008
믿을 수 없는 - 믿을 수 있는	.656	.039	.278
어려운 - 쉬운	.000	.749	-.027
흥미롭지 않은	.319	.635	.039
불성실한 - 성실한	.551	-.180	.256
필요하지 않은 - 필요한	.710	.103	.242
위험한 - 안전한	.291	.232	.617
무거운 - 가벼운	-.008	.671	.000
익숙하지 않은 - 익숙한	.090	.562	.162
도움이 되지 않는 - 도움이 되는	.615	.088	.330
부담이 되는 - 부담 안 되는	.047	.746	-.065
더러운 - 깨끗한	.196	-.007	.814
나쁜 - 좋은	.352	.136	.711
옳지 않은 - 옳은	.468	.115	.601
가치 없는 - 가치 있는	.662	.035	.399
재미없는 - 재미있는	.190	.720	.129
불완전한 - 완전한	.507	.069	.465
고맙지 않은 - 고마운	.473	.414	.382
쓸모없는 - 쓸모 있는	.601	.092	.477
중요하지 않은 - 중요한	.630	.036	.338
즐겁지 않은- 즐거운	.318	.695	.142

〈표 3〉 측정도구의 신뢰도

차원	형용사(문항) 수	Cronbach α
활동	12	.904
능력	10	.881
평가(evaluative factor)	4	.814

대학생은 2018년 1학기 교직이론 「교육평가」 과목을 수강한 E 사범대학과 교직과정 학생들과 S대학 교직과정 학생들이며, 이와 함께 초·중등 현직 교사들을 대상으로 하였다. 구체적인 집단별 구성은 <표 4>와 같다.

〈표 4〉 연구대상

구분	대상	빈도	퍼센트	비고
‘평가’ 이해관련자	초등학생(5~6학년)	184	30.46	프로파일 분석 및 ANOVA
	중학생(1학년)	101	16.72	
	고등학생(1학년)	97	16.06	
	대학생(교육평가수강자)	115	19.04	
	초·중등학교 교사	107	17.72	
	합계	604	100	

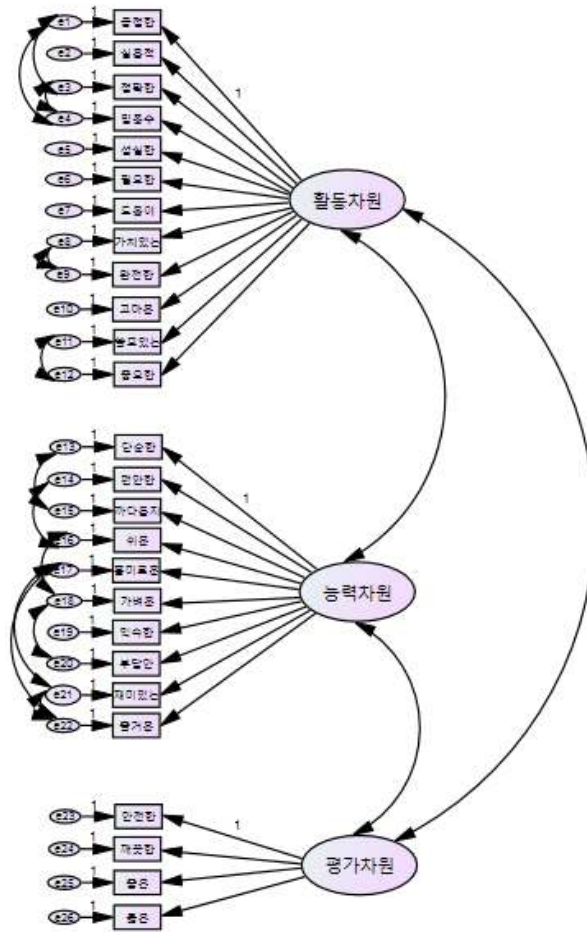
2. 분석방법

1) 평가 의미 분석

의미 분석 척도에서 변산을 초래하는 원천은 크게 3가지(개념, 척도, 피검사자)이지만(권순달, 2013), 본 연구에서는 척도 간, 피검사자간의 변산을 중심으로 분석하였다. 척도간 분석이란 3 차원을 표현하는 형용사군의 평균값을 산출하였으며, 이를 바탕으로 프로파일 분석을 하였다. 이와 함께 집단별 ANOVA방법으로 분석하였으며, 차이가 있을 경우 Scheffe검정으로 사후분석을 실시하였다.

2) 평가 의미 구조 분석

‘평가’ 의미 확인적 구조 분석을 위하여 ‘평가’의 3 차원, 활동 차원 12개 형용사, 능력 차원 형용사 10개, 평가차원(evaluative factor) 4개 형용사를 측정변수로 하여 [그림 5]와 같이 구조방정식모형으로 분석하였다.



(그림 5) 평가 의미 구조 모형

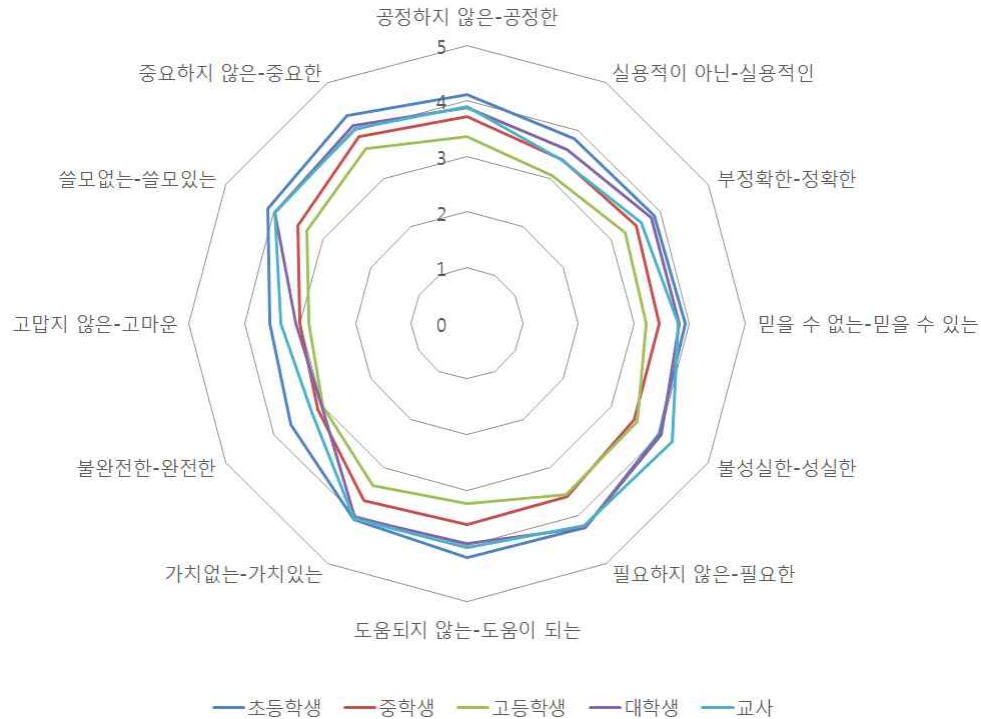
IV. 연구 결과

1. 프로파일 및 차이 분석

1) 활동차원

‘평가’ 의미 3차원 중 ‘활동차원’ 12개 형용사에 대한 집단별 프로파일은 [그림 6]과 같다.

[그림 6]에 의하면 ‘활동차원’에서 중학생과 고등학생이 초등학생과 비교하여 각 형용사에 대하여 부정적인 인식을 지니고 있는 것으로 나타났다. 한편, ‘평가’ 활동차원에서 불완전하



(그림 6) '평가'의 활동차원 프로파일

고, 고맙지 않은 것에서 점수가 낮게 나타났다.

한편 위의 '활동차원' 각 형용사에 대한 집단 차이를 비교하기 위하여 분산분석한 결과는 <표 5>와 같다.

<표 5>에서 초등학생의 점수가 다른 집단보다 높게 나타났으며, 그러한 차이는 유의한 것으로 밝혀졌다. 다만 학생들과 달리 교사들은 '평가'를 성실하게 수행하는 것으로 나타났다.

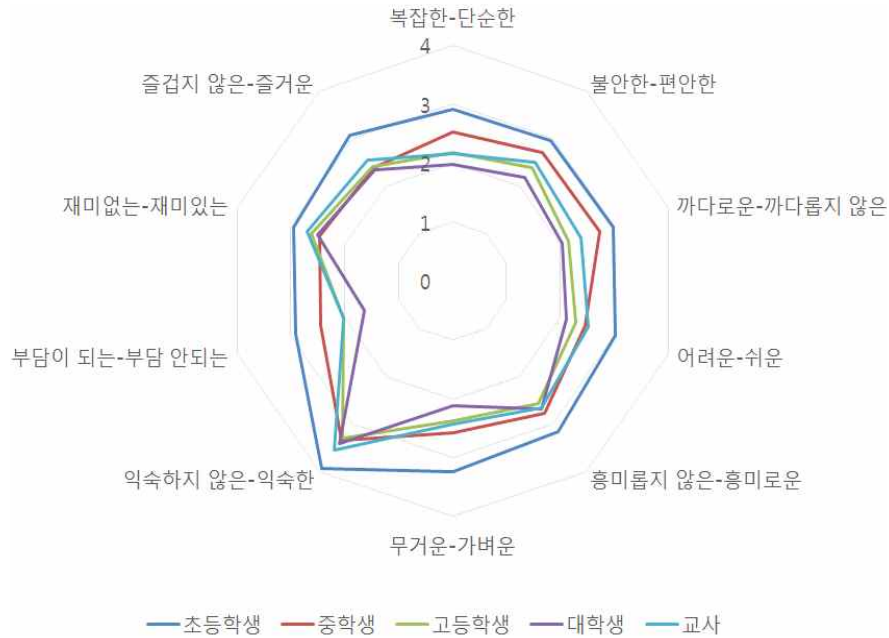
2) 능력차원

'평가' 의미 3차원 중 '능력차원' 10개 형용사에 대한 프로파일은 [그림 7]과 같다.

[그림 7]에 의하면 '능력차원'에 대한 형용사에서 대부분 부정적인 형용사에 응답을 하였다. '부담이 되는 - 부담 안 되는' 형용사에서 '부담이 되는 것'으로 응답하였으며, '까다로운 - 까다롭지 않은' 형용사에서도 '까다롭다'고 응답했으며, '어려운 - 어렵지 않은' 형용사에서 '어렵다'고 응답하였다. 그리고 복잡하고, 흥미롭지도 않으며, 즐겁지 않고, 불안하다고 응답하였다.

〈표 5〉 '평가'의 활동치원 분산분석표

형용사	집단(명)												F	사후검정		
	초(84)			중(101)			고(97)			대(115)						
	M	SD		M	SD		M	SD		M	SD					
공감하지 않은-공감한	4.12	1.006		3.71	1.042		3.36	1.174		3.88	.818		3.90	.980	9.609	초=교=대>고
실용적인 아닌-실용적인	3.84	1.006		3.40	1.087		3.07	1.166		3.61	.943		3.39	1.062	9.637	초>중=교, 대>고
부정확한-정확한	3.88	.975		3.51	1.045		3.27	1.066		3.81	.634		3.62	.997	7.792	초>고, 대>고
믿을 수 없는-믿을 수 있는	3.92	1.079		3.45	1.044		3.21	1.03		3.81	.620		3.80	.840	11.540	초>고, 초>중, 교>고
불성실한 -성실한	3.98	.932		3.46	1.044		3.53	1.062		4.02	.562		4.26	.850	15.351	초>중=고, 대>중=고, 교>중=고
필요하지 않은-필요한	4.25	1.036		3.59	1.21		3.56	1.099		4.24	.696		4.21	.774	15.361	초>중=고, 대>중=고, 교>중=고
도움이 되지 않는-도움이 되는	4.22	1.006		3.62	1.165		3.24	1.125		3.96	.754		4.04	.835	18.292	초>고, 교>고, 교>중=고
가치 없는-가치 있는	4.08	1.016		3.69	1.056		3.38	1.045		4.04	.821		4.06	.750	11.343	초=교=대>고
불완전한-완전한	3.66	.989		3.09	1.123		2.99	.963		3.01	.874		3.21	1.007	12.076	초>중=고=대=교
고맙지 않은-고마운	3.55	1.12		3.00	1.183		2.84	1.17		3.07	.722		3.34	.890	10.199	초>고, 교>고, 초>중=대
쓸모없는-쓸모있는	4.14	1.05		3.52	1.221		3.32	1.114		3.99	.789		3.98	.777	14.149	초=대=교>중=고
중요하지 않은-중요한	4.32	.991		3.87	1.163		3.62	1.075		4.11	.770		4.04	.910	8.927	초>고, 대>고, 교>고



(그림 7) '평가'의 능력차원 프로파일

한편 위의 '능력차원' 각 형용사에 대한 집단 차이를 비교하기 위하여 분산분석한 결과는 <표 6>과 같다.

<표 6>에서 전체적으로 대학생들이 다른 집단보다 낮게 나타났으며, '평가'에 '익숙한 것'으로 밝혀졌다. 집단별 점수 차이는 유의미한 것으로 나타났다.

3) 평가차원(evaluative factor)

'평가' 의미 3차원 중 '평가차원' 4개 형용사에 대한 프로파일은 다음 [그림 8]과 같다.

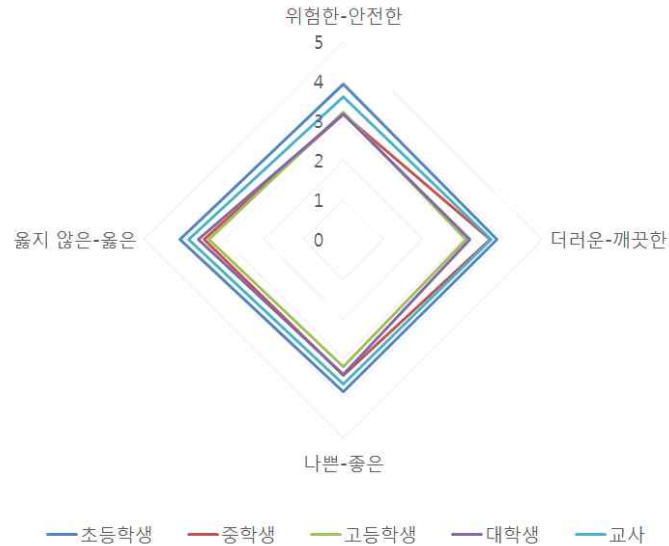
[그림 8]에 의하면 '평가차원'에 대한 형용사에서 '수동적 - 능동적'에서 수동적으로 응답 하였으며, 다른 형용사에서는 대부분 보통(3점) 이상으로 나타났으며, 특히 '평가'는 매우 섬세하게 이루어지고 있는 것으로 인식하고 있음을 알 수 있다.

한편 위의 '평가차원(evaluative factor)' 각 형용사에 대한 집단 차이를 비교하기 위하여 분산분석한 결과는 <표 7>과 같다.

<표 7>에서 모든 집단에서 '평가'는 안전하고, 깨끗하며, 좋으며, 옳은 것으로 판단하고 있는 것으로 밝혀졌다. 특히 초등학교에 비하여 중고등학교 학생들은 낮은 점수를 보여주고 있다. 집단별 점수 차이는 유의미한 것으로 나타났다.

〈표 6〉 ‘평가’의 능력치원 분산분석표

형용사	집단(명)										F	사후검정
	초(84)		중(101)		고(97)		대(115)		교사(107)			
	M	SD	M	SD	M	SD	M	SD	M	SD		
복잡한-단순한	2.92	1.050	2.53	1.064	2.18	1.127	1.98	.749	2.16	.754	21.922	초>대=교사=고, 중>대
불안한-편안한	2.95	1.275	2.70	1.136	2.38	1.220	2.17	.787	2.49	.925	10.215	초>고>,대 중>대, 중>고
까다로운-까다롭지 않은	2.98	1.250	2.73	1.165	2.16	1.017	2.04	.831	2.38	.948	17.998	초>대, 초>교, 초=중>고
어려운-쉬운	3.01	1.254	2.46	1.171	2.29	1.050	2.11	.672	2.51	.873	15.507	초>중=고=대=교사
흥미롭지 않은-흥미로운	3.16	1.291	2.77	1.240	2.58	1.273	2.68	1.072	2.66	1.205	5.414	초>고
무거운-가벼운	3.23	1.264	2.57	1.143	2.37	1.064	2.11	.698	2.43	.912	23.872	초>대, 중>대
익숙하지 않은-익숙한	3.93	1.082	3.34	1.227	3.31	1.270	3.41	1.091	3.56	1.142	7.331	초=교>중=고=대
부담이 되는-부담 안 되는	2.91	1.360	2.45	1.277	2.01	1.177	1.63	.717	2.01	1.032	25.469	초=중>고=교>대, 초>중
재미없는-재미있는	2.95	1.186	2.46	1.196	2.62	1.211	2.51	.949	2.70	1.011	4.314	초>중
즐겁지 않은-즐거운	3.07	1.297	2.38	1.224	2.40	1.296	2.33	.845	2.54	1.058	10.647	초>중,초>고,초>대, 초>교



(그림 8) '평가'의 평가차원(evaluative factor) 프로파일

4) 3차원 비교

‘평가’ 의미의 3 차원(활동차원, 능력차원, 평가차원) 형용사의 평균 점수를 종합하여 공간적으로 그 거리를 나타내면 [그림 9]와 같다.

위의 [그림 9]에서 ‘평가’의 의미 3 차원의 점수는 평가차원(evaluative factor)(3.45), 능력차원(3.32), 활동차원(2.30)의 순으로 나타났다. 한편, ‘평가’ 의미의 3 차원(활동차원, 능력차원, 평가차원)의 평균 점수를 대응표본 t검정한 결과는 <표 8>과 같다.

<표 8>에 의하면 3차원에서 능력차원의 점수가 가장 낮으며 다음으로 평가차원, 활동차원의 순으로 나타났다. 3차원(활동차원, 능력차원, 평가차원)의 평균점수 차이는 유의미한 것으로 나타났다.

2. 구조방정식모형 분석

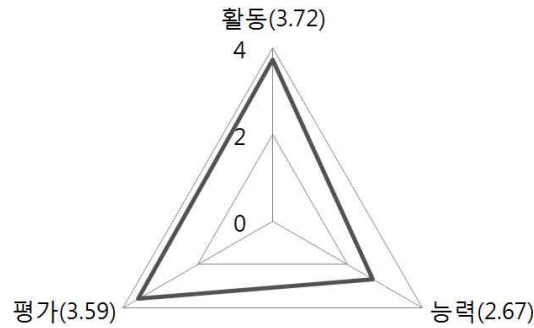
1) 확인적 요인분석

‘평가’의 3개 차원 26문항의 요인계수-비표준화 회귀계수, 표준화 회귀계수, 표준오차-는 <표 9>와 같다.

<표 9>에 의하면 표준화 회귀계수는 .899에서 1.699 사이의 값을 나타내고 있으며, 모든 계수는 유의미한 것으로 나타났다. 이는 문항(측정변수)들이 3개 차원(활동차원, 능력차원, 평가차원)을 타당하게 측정하고 있음을 말해준다. 따라서 <표 9>에서 형용사 문항들이 적

〈표 7〉 '평가'의 평가차원(evaluative factor) 분산분석표

형용사	집단(명)										F	사후검정
	초(840)		중(101)		고(97)		대(115)		교사(107)			
	M	SD	M	SD	M	SD	M	SD	M	SD		
위험한-안전한	3.93	1.104	3.17	1.096	3.21	1.030	3.17	.794	3.61	.888	17.019	초=교사>중=대 교사>중=대 초>고
더러운-깨끗한	3.85	.991	3.69	1.056	3.07	1.023	3.19	.620	3.71	.789	17.341	중=교=초>고=대
나쁜-좋은	3.83	1.013	3.42	1.160	3.21	1.030	3.40	.698	3.63	.841	8.379	초=교>중=대>고
옳지 않은-옳은	4.10	.959	3.51	1.029	3.39	.963	3.66	.661	3.90	.752	13.740	초>대>중=고 교>중=고



[그림 9] '평가'의 3차원 프로파일

<표 8> '평가'의 3 차원 검정 결과

대응	구분	평균	N	표준편차	t	p
활동차원		3.7275	604	.70965	35.338	.000
능력차원		2.6682	604	.81147		
활동차원		3.7275	604	.70965	6.612	.000
평가차원		3.5861	604	.78691		
능력차원		2.6682	604	.81147	-31.710	.000
평가차원		3.5861	604	.78691		

합하게 측정하고 있는 것으로 밝혀져, 다음 단계로 본 연구모형의 적합도를 검정하였으며, 결과는 [그림 9]와 같다.

[그림 10]을 보면, 일부 측정 항목들 사이의 공분산이 존재하고 있다. 이는 [그림 2] 연구 모형에서 제시했던 바, 일부 항목들의 진술이 중복되어 있음을 의미한다. 3개 차원(이론 변수) 사이의 공분산과 유의도 검정 결과는 <표 10>과 같다.

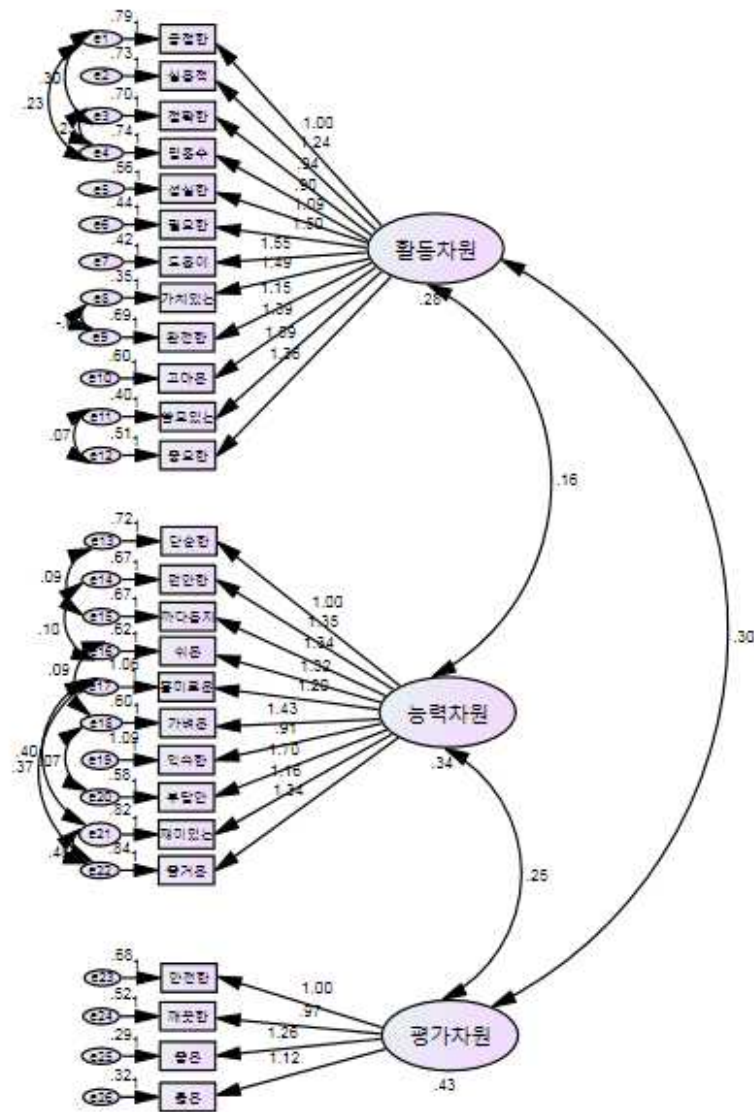
<표 10>에 의하면 공분산의 값은 모두 유의미한 것으로 나타났다. 물론 오차변수 사이의 공분산을 자유화시킴으로써 적합도를 높일 수 있지만, 그렇다고 논리적으로 설명이 불가능한 상태에서 오차변수 사이의 상관을 가정할 수는 없었다. 항목들의 의미를 고려하여 오차변수의 공분산의 존재를 고려할 때, 모형의 적합도 지수는 <표 11>과 같이 나타났다.

일반적으로 SEM의 적합도 지수로 χ^2 , RMSEA, GFI, NNFI, SRMR등이 사용되고 있다(문수백, 2015). 한편, RMSEA, TLI(NNFI), CFI, GFI 등으로 충분하다는 주장도 있다(성태제, 2015).

〈표 9〉 요인계수

			Estimate	S.E.	C.R.	P
필요한	<---	활동차원	1.496	.120	12.462	***
도움이	<---	활동차원	1.550	.123	12.597	***
가치있는	<---	활동차원	1.486	.117	12.668	***
완전한	<---	활동차원	1.149	.107	10.774	***
고마운	<---	활동차원	1.393	.118	11.793	***
쓸모있는	<---	활동차원	1.588	.125	12.684	***
중요한	<---	활동차원	1.361	.114	13.971	***
성실한	<---	활동차원	1.095	.099	11.037	***
믿을수	<---	활동차원	.899	.079	11.370	***
정확한	<---	활동차원	.943	.073	12.834	***
실용적	<---	활동차원	1.243	.113	11.026	***
공정한	<---	활동차원	1.000			
흥미로운	<---	능력차원	1.195	.110	10.895	***
가벼운	<---	능력차원	1.426	.112	12.746	***
익숙한	<---	능력차원	.913	.098	9.305	***
부담안	<---	능력차원	1.699	.126	13.476	***
재미있는	<---	능력차원	1.162	.102	11.407	***
즐거운	<---	능력차원	1.344	.111	12.073	***
쉬운	<---	능력차원	1.322	.105	12.535	***
까다롭지	<---	능력차원	1.344	.099	13.518	***
편안한	<---	능력차원	1.353	.108	12.543	***
단순한	<---	능력차원	1.000			
깨끗한	<---	평가차원	.971	.072	13.520	***
좋은	<---	평가차원	1.263	.079	16.042	***
옳은	<---	평가차원	1.122	.072	15.489	***
안전한	<---	평가차원	1.000			

<표 11>에서 보면 유의확률이 .000으로 모형이 적합하지 않다고 할 수 있으나, χ^2 통계량은 표본의 수에 민감하게 영향을 받을 수 있기 때문에 다른 적합도 지수로 판단할 필요



(그림 10) '평가' 의미 구조방정식모형 분석 결과

〈표 10〉 3개 차원의 공분산과 유의도

			Estimate	S.E.	C.R.	P
능력차원	<-->	활동차원	.156	.021	7.314	.00
능력차원	<-->	평가차원	.245	.029	8.538	.00
활동차원	<-->	평가차원	.302	.033	9.094	.00

〈표 11〉 모형의 적합도 지수

	χ^2	df	p	RMSEA	GFI	NNFI	CFI	SRMR
연구모형	1102.417	284	.000	.069	.874	.872	.901	.091

가 있다. 전체적으로 적합도 지수는 크게 만족스럽지 못하다. 일반적으로 GFI, NNFI, CFI 지수는 .90 이상이면 적합하다고 할 수 있는데, 일부 .90보다 낮게 나타났으며, RMSEA는 .05이하, SRMR은 .08이하로 나타나면 적합도가 매우 좋은 것으로 판단할 수 있는데(배병렬, 2009), 이러한 지수들이 조금 높게 나타났다. 따라서 본 연구 모형은 향후 수정될 필요가 있다. 그러나 이론적 배경을 무시하고, 적합도 지수를 높이기 위하여 측정변수들의 오차공분산을 가정하는 것 바람직하지 않기 때문이다(배병렬, 2009).

V. 결 론

1. 논의

본 연구는 ‘평가’ 개념 이미지를 분석하는데 있다. 이를 위하여 Osgood이 제안한 의미분석 방법에 따라 양극을 이루는 형용사를 사용하였다. 26 개 형용사를 사용하여 요인분석 결과 활동차원 12개, 능력차원 10개, 평가차원 4개, 총 26개의 형용사를 사용하였다. ‘평가’ 관련 이해관련자(stakeholder)라고 생각되는 경기도교육청과 서울특별시교육청 관내 초등학교, 중학생, 고등학교, 대학생, 및 교사 604명을 대상으로 2018년 4월 5일부터 5월 4일까지 자료를 수집하였다.

연구방법으로 ‘평가’의 3차원을 나타내는 형용사들을 집단별로 프로파일 분석과 각 형용사에 대하여 집단별 ANOVA를 실시하고, 차이가 있을 경우 사후검정을 하였다. 그리고 26개 문항의 3차원 구조관계를 파악하기 위하여 구조방정식모형으로 적합도를 분석하였다.

프로파일 분석 결과, ‘평가’의 의미 3차원(활동차원, 능력차원, 평가차원)형용사들은 차이가 있는 것으로 나타났다. 활동차원을 의미하는 12개의 형용사들에서 5점 척도에서 3점 척도 이상으로 가장 높게 나타나, ‘평가’의 실제 활동에 대해서는 학생이나 교사 모두 긍정적으로 인식하고 있는 것으로 나타났다. ‘평가’는 필요하고, 도움이 되고, 가치 있으며, 쓸모 있고, 중요하며, 공정하게 이루어지고 있다고 인식하고 있는 것으로 밝혀졌다.

그러나 능력차원의 형용사에 대한 응답은 대부분 3점(보통)이하로 가장 낮게 나타났다.

‘평가’에 익숙하지 않으며, 불안하고, 흥미롭지 않고, 재미없고, 즐겁지 않은 것으로 나타났다. 활동 차원에서 ‘평가’는 매우 까다롭고, 매우 부담이 되며, 어려운 것으로 밝혀졌다. 이러한 결과는 ‘평가’ 자체에 대한 부정적인 인식이 원인이라고 할 수 있다.

평가차원에서는 공정하고, 꼼꼼해야 하고, 정확해야 하는 것으로 인식하고 있음을 알 수 있다. 따라서 3 차원- 능력차원, 활동차원, 평가차원-의 점수의 차이분석에서 유의미한 것으로 나타났다. 그리고 3차원 점수에서 초등학생들의 점수가 다른 집단보다 높게 나타났다. 특히 ‘평가’의 활동차원에서 ‘평가’는 불완전하며, 별로 고맙지 않다고 여겨 가장 낮은 점수를 보여주고 있다. 능력차원에서는 ‘평가’에 익숙하다는 것에 가장 높게 나타나, 학교 현장에서 ‘평가’가 일상적으로 수행되고 있음을 알 수 있다. 평가차원에서는 중학생들이 다른 집단보다 전체적으로 낮은 점수를 보였다. 특히 ‘평가’ 의미의 능력차원 각 형용사에 매우 낮은 점수를 나타낸 것은 실제 ‘평가’를 수행하는 현장에서 평가전문가나 교사에게 시사하는 바가 크다. 교육 ‘평가’ 수업을 통하여 ‘평가’의 의미를 통해 긍정적인 생각을 가지도록 할 필요가 있다.

한편 구조방정식으로 검정할 결과, 형용사의 측정변수들의 요인계수들이 유의한 것으로 나타났으나, 모형의 적합도에서 카이스퀘어 값은 유의하지 않았으나, 다른 적합도 지수는 기준을 충족하고 있었다. 다만, 후속 연구를 통하여 유사한 형용사들을 통합하여 측정변수의 수를 줄인다면 모형의 적합도는 다소 높아질 것으로 판단된다.

2. 제언

교육에서 ‘평가’는 매우 중요한 기능을 하고 있다. ‘평가’ 태도나 ‘평가’ 유형의 연구가 아니라, ‘평가’ 그 자체에 대한 의미 분석을 수행하면서, 이를 형용사로 표현하는 것이 쉽지 않았다. 그 이유는 ‘평가’ 그 자체가 추상명사이기 때문이다. 그러나 ‘평가’에 대한 의미 분석은 ‘평가’ 전문가에게 많은 시사점을 주고 있다. ‘평가’를 가르치는 사람은 ‘평가’ 그 자체에 대하여 긍정적인 인식을 심어줄 필요가 있으며, 교사를 비롯한 실제 ‘평가’ 담당자는 평가활동에서 더욱 공정하고 믿음을 가지게 할 필요가 있다.

그리고 3개 차원(활동차원, 능력차원, 평가차원)의 형용사를 ‘평가’의 개념에 적용하는데 어려움이 있음을 본 연구를 통해 발견할 수 있었다. 기존에 소개되고 있는 의미분석에 사용되는 형용사들은 일반명사를 표현하는데 적절하기 때문이다. 그 이유는 형용사란 사물의 성질이나 상태를 나타내는 품사이기 때문에 ‘평가’와 같은 추상 명사를 형용사로 표현하는 것이 적절하지 못할 수 있기 때문이다. 즉, 3차원으로 분류된 형용사들이 실제 탐색적 요인분석 결과 일부 다르게 나타났기 때문이다. 그리고 형용사를 단순히 양극으로 나타내는 것

이 적절한 것인가에 대해 심층적인 연구가 필요할 것이다. 그럼에도 불구하고 ‘평가’에 대한 정의적 특성(인식, 태도, 감정 등)이 아닌 ‘평가’의 순수한 의미를 분석하는 것은 의미가 있을 것이다. 이를 바탕으로 ‘평가’ 수업을 받은 집단과 그렇지 않은 집단에 대한 ‘평가’ 의미 비교도 가능할 것이다. 그리고 분석 방법으로 단순한 프로파일이 아닌 Osgood이 제안한 D통계치를 사용하여 심층적으로 분석한다면 보다 유용한 결과를 얻을 수 있을 것이다.

참고문헌

- 권순달(2018). 평가 의미 분석. **한국교육평가학회 2018춘계학술대회 자료집**, 1-15.
- 권순달(2013). 평가의 오용과 남용. **교육평가연구**, 26(4), 739-754.
- 권순달(2011). **교육연구의 이해**. 경기: 양서원.
- 권순달·오성삼(2016). **교육평가**. 경기: 양서원.
- 노현중(2017). 평가에 대한 태도 척도(STAT-Q) 개발 및 타당화: 총괄 및 형성평가에 따른 비교. 경북대학교 대학원 박사학위논문.
- 노현중·손원숙(2018). 초등학생용 평가에 대한 태도 척도(SATA-Q)개발 및 타당화. **교육평가연구**, 31(1), 155-179.
- 문수백(2015). **구조방정식모델링의 이해와 적용**. 서울: 학지사.
- 박영례(2017). 유이 인성리더십 척도 개발 및 타당화. 숙명여자대학교 대학교 박사학위논문.
- 배병렬(2009). **Amos 17.0 구조방정식모델링-원리와 실제**. 서울: 청람.
- 성태제(2016). **교육연구방법의 이해**. 서울: 학지사.
- 성태제(2015). **알기 쉬운 통계분석**. 서울: 학지사.
- 성태제(2014). **현대교육평가**. 서울: 학지사.
- 송영옥·최혁준(2017). 의미분석법에 의한 물리적 이미지 측정도구 개발 및 적용. **한국과학교육학회지**, 37(6), 1051-1061.
- 윤희성(2016). 의미분석법에 의한 초등학생들의 로봇에 대한 이미지 연구. 광주교육대학교 교육대학원 석사학위논문.
- 이상미(2016). 의미분석법을 이용한 융합과학 이수 고등학생들의 과학 영역별 이미지 변화: 화학을 중심으로. 한국교원대학교 대학원 석사학위논문.
- 이순목(1995). **요인분석 I**. 서울: 학지사.
- 이영만(2015). 조직공정성, 인사평가인식, 조직문화가 조직효과성에 미치는 영향: 경기도 교육행정관을 중심으로. 경기대 대학원 박사학위논문.

- 이영만·하봉운(2015). 조직공정성, 인사평가인식, 조직문화가 조직효과성에 미치는 영향 분석. **교육행정학연구**, 33(1), 129-150.
- 이재완(2017). 마을만들기 사업의 참여경험이 평가인식에 미치는 영향. **한국지역사회복지학**, 61(-), 29-51.
- 이종승(1984). **교육연구법**. 서울: 배영사.
- 이종찬(2009). 상사의 배려행동, 상사-부하 간 교환관계가 조직시민행동 및 성과 평가인식에 미치는 영향. **대한경영학회지**, 22(5), 2979-2998.
- 조영민(2017). 고등학생의 평가인식 유형 분석: 성취목표지향성, 시험불안 및 학업참여와의 관계. 경북대학교 교육대학원 석사학위논문.
- 조영민·손원숙(2018). 교실평가에 대한 고등학생의 인식 유형 분석. **교육과정평가연구**, 21(1), 197-217.
- 조운교(2010). **의미분석법에 의한 초등학생의 친구 이미지 분석**. 대구교육대학교 교육대학원 석사학위논문.
- 최종석(2014). 의미분석법을 활용한 불교이미지에 관한 연구-종립대학 대학생들을 중심으로. **한국불교학**, 69, 419-444.
- 표경현·나경희(2013). EFL 학습자의 자기평가에 대한 인식 변화 분석: 사회문적 관점에서. **교과교육학연구**, 17(4), 1169-1194.
- 한국교육평가학회(2005). **교육평가 용어사전**. 서울: 학지사.
- 홍소영(2018). 학생 자기평가의 학습효과에 관한 메타분석. **교육평가연구**, 31(1), 309-331.
- 황정규(2002). **학교학습과 교육평가**. 서울: 교육과학사.
- www.natmal.com
- krdic.naver.com
- <http://tutorials.istudy.psu.edu/testing/testing2.html>
- <https://en.wikipedia.org/wiki/Evaluation>
- <https://en.oxforddictionaries.com/definition/evaluation>
- <http://journals.sagepub.com/home/evi>
- <https://dictionary.cambridge.org>
- Ding, Z., & Ng, F. (2008). A new way of developing semantic differential scales with personal construct theory. *Construction Management and Economics*, 26, 1213-1226.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2003). *Program Evaluation: Alternative Approaches and Practical Guidelines*. 3d ed. Boston: Allyn and Bacon.
- House, E. R. (1991). Evaluation and Social Justice: Where are we? in Milbery W. McLaughlin & D.

- C. Phillips(ed). Evaluation and Education: At Quarter Century. 233-247. Chicago; The University of Chicago press.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). Evaluation Theory, Models, and Applications. CA: Jossey-Bass.

© 논문접수: 2018. 7. 31 / 수정본 접수: 2018. 9. 14 / 게재승인: 2018. 9. 15

— 저 자 소 개 —

· 권순달 : 서울교육대학교 졸업. 고려대학교 교육학석사, 교육학박사. 교육프로그램 평가 전공.
현재 수원대학교 교수로 재직 중임. 관심 분야는 프로그램 평가 모형 개발과 검증.
sdkwon@suwon.ac.kr

〈ABSTRACT〉

Evaluation Semantic Differential of Stakeholders

Soon-Dal, Kwon

The University of Suwon

This study is to analyze the meaning of stakeholders in 'evaluation'. To do this, semantic differential methods were used. For this purpose, the adjectives representing the three dimensions (ability, activity, and evaluation) of the concept of 'evaluation' were selected and the logical validity of the teachers was tested. With 12 activity dimension adjectives, 10 ability dimension adjectives, and 4 evaluation dimension adjectives are selected 26 adjectives. 604 subjects were collected and exploratory factor analysis (EFA) was performed. Profile analysis and ANOVA were conducted. Structural equation modeling were used to analyze structural relationships. As a result of this study, the score of activity dimension was the highest among the three dimensions, and both middle and high school students and teachers except the elementary students recognized that they were negative in the ability dimension. Therefore, it is necessary to have a positive perception of the actual evaluation, and it is necessary to raise fairness in the evaluation activities. These results are highly suggestive to teachers in charge of teaching evaluation.

Keywords : Evaluation, Semantic differential, Structure Equation Modeling, EFA