

학업 성취도 평가도구의 내용 타당도 분석

- 수학과 3-가를 중심으로 -

김은아¹⁾ · 강완²⁾

본 연구의 목적은 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구와 교사 제작 학업 성취도 평가 도구의 내용타당도를 분석·비교하는데 있으며, 학교 현장에서 실시되는 학업 성취도 평가 도구의 개선을 위한 기초 자료를 제시하는데 그 의의가 있다. 학업 성취도 평가 도구의 내용 타당도를 검증하기 위하여 행동차원별, 내용차원별로 수업목표와 평가목표의 비율차 검증을 하였다. 그 결과 교사 제작 수학과 3-가 학업 성취도 평가도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가도구는 행동차원과 내용차원에서 모두 내용타당도가 높은 것으로 나타났다. 이 연구는 교사 제작 학업 성취도 평가도구는 내용타당도는 높으나 평가 문항이 지식, 이해력, 적용력을 평가하는데 집중된 경향을 보여 창의력 신장을 위한 고등정신기능을 기르는 평가가 이루어지도록 평가 방법 개선 방안과 평가 도구 개발에 힘써야 할 것이며 각 학교간의 편차를 해소하기 위하여 많은 교사들의 상호 교류도 활발히 진행되어야 할 것임을 제언한다.

[주제어] 교육평가, 학업 성취도 평가도구, 타당도, 내용타당도

I. 서 론

사람이 살아가는 일은 단순히 목숨을 유지하는 일만이 아니라, 인간 삶의 질을 개선시키고자 하는 지속적인 노력이라 할 수 있다. 교육에서의 평가도 이와 같은 의미를 갖는 것이라고 보아야 할 것이다. 우리는 교육평가에 대한 학습의 출발로서 '무엇을 어떻게 평가할 것인가?'라는 질문을 탐구하기 전에 먼저 '왜 평가하는가?' 라는 본질적인 질문을 생각해 볼 필요가 있다. 이것은 곧 교육평가란 무엇이며 어떠한 성질의 활동인가를 이해하는 일이 될 것이다.

교육평가는 교육목표의 달성 정도를 확인하는 학교활동으로서 학습자의 행동변화 및 학습과정에 관한 정보를 수집하고 그 결과를 학습의 극대화를 위해 이용할 수 있다는 관점에서 볼 때 학교교육에서 필수적이다. R. W. Tyler(1949)는 "평가의 과정이란 본질적으로 교육과정 및 수업의 프로그램에 의하여 교육 목표가 실제로 어느 정도나 실현되었는지 밝히는 과정이다."라고 정의했다. 이는 어떤 교육과정에 있어서 그것을 달성하고자 하는 교육목표가 내포되어 있음을 의미한다.

교육평가의 본질적 기능은 학습자를 등급별로 나누어 서열화하는 것이 아니라 교수·학

1) [제1저자] 서울 정심초등학교

2) [교신저자] 서울교육대학교 수학교육과

습 과정에서 학습자가 해 나가는 학업 수행 상황을 진단하고, 학업 성취 정도를 판단하는데 필요한 자료를 제공하는 것이다. (교육인적자원부, 2006) 학교 교육에서 교육평가의 중요한 역할 중의 하나는 학생 개개인의 학업 성취도를 평가하는 일이다. 학업성취도의 평가는 학습 결과 뿐 아니라 학습자의 학업 수행 과정과 상황을 진단하고 학업 성취 정도를 판단하는데 필요한 자료를 제공하는 것이며 내용의 습득 정도, 목표 달성 정도, 개인의 성장 발달 정도 등을 파악하여 지도 자료로 활용함으로써 학생들을 바람직한 방향으로 이끌어 가도록 하는 것이다. 이러한 교육평가는 교육과 관련된 모든 현상을 분석하여 교수·학습 효과를 극대화시키기 위하여 도와주는 기능을 하여야 하며 타당한 평가도구를 필요로 한다. 목표가 명확히 정의되어 있느냐는 교육에 의한 학습자의 의도된 변화를 평가하기 위한 전제조건이다. 이에 따라 교육평가에서의 중요한 관심사는 교육목표 달성 정도를 측정할 수 있는 평가도구의 신뢰도나 객관도 보다는 타당도, 특히 내용타당도(content validity)에 관한 것이다. 이는 타당도가 높은 평가도구에 의한 교육평가가 이루어졌을 때 교육평가 의미와 가치를 찾을 수 있기 때문이다.

이처럼 평가의 중요성이 강화되고 있으며 이에 따라 최근 들어 우리 학교 교육현장은 교사에게 수준 높은 학생평가 전문성을 요구하고 있으며 다양한 교사 연수를 통하여 개선하려고 노력하고 있다. 따라서 본 연구의 목적은 KICE와 초등학교 교사들이 제작한 학업성취도 평가도구의 내용타당도를 분석함으로써 학업성취도 평가도구가 학생들의 학업성취도를 측정하는 객관적인 잣대로서의 기능을 충분히 하고 있는지를 알아보고자 하며 학교에서 활용하는 교사 제작 학업 성취도 평가 도구의 신뢰성을 높이고 보다 전문성을 가지는 평가 도구의 개선을 위한 기초 자료로 제시하는데 그 의의가 있다.

II. 이론적 배경

1. 교육평가의 개념

교육평가는 교수·학습 활동의 목적과 내용 및 방법, 그리고 결과에 대한 체계적인 정보를 제공해주는 것이다. 교육현상의 가치나 질에 대한 판단을 내리는 것으로 보기도 하고 교육 가치를 실현하는 것을 교육평가로 규정하기도 한다(채선희, 1998). 교육평가는 학교현장에서는 흔히 검사, 시험 등으로 널리 알려져 있는 용어이며 교육평가의 개념과 유사하나 교육평가란 용어와 완전히 같은 의미는 아니다. 교육평가의 개념 속에 검사나 시험이라는 용어가 포함된다고 볼 수 있다. 그만큼 교육평가의 개념은 매우 광범위한 의미를 담고 있으며 교육평가의 의미를 어떻게 정의하느냐에 따라 평가의 목적이 달라진다.

교육평가의 목적은 교육을 구속하는 기능이 아니라 교육을 도와주는 기능이므로 교육평가가 교육대상에 긍정적 영향을 주어야 한다. 교육평가를 교육프로그램에 대한 의사결정을 위하여 정보를 사용하거나 수집하는 과정이라고 정의하면 평가 대상의 가능한 정보를 수집하여 교육적 의사결정을 하거나 이를 도와주는 기능을 평가의 목적이라 할 수 있다. 그러므로 교육평가는 학습과 교육과정에 최대한 도움을 주어 학습을 극대화시키려는 데 목적이 있다(성태제, 2002). 교육평가는 교육과정을 평가하고 교육 자료나 교육 프로그램을 개선하거나 학업 성취 수준을 총평하는 등 다양한 목적을 지니고 있다. 목적이 있는 곳에는 그것을 달성하기 위한 기능이 수반되기 때문에 교육평가의 의미를 어떻게 정의하느냐에 따라 평가의 목적이 달라지고 평가의 목적이 다르면 평가의 기능이 달라진다.

교육평가의 절차에는 평가하려는 대상에 따라 어느 정도의 차이가 있으며 지능, 학력, 성격, 흥미, 환경 등의 어느 것을 평가하느냐에 따라 평가의 방법과 절차 도구 면에서 다를 수 있다. 그러므로 정확한 평가와 더불어 평가대상과 평가도구를 선정, 제작하는 일은 매우 중요한 교육평가의 실체라고 할 수 있다. 교육평가의 기본 절차에 대한 논의는 목적과 방법에 따라 다양하다. 수업목표 확인, 평가 장면의 선정, 평가도구의 제작 또는 선정, 평가의 실시 및 결과 처리, 평가결과의 해석 및 활용이라는 5단계의 절차로 구분한다(김대현, 김석우, 2005). 변창진 외(1996)는 교육평가의 기본 절차를 평가목적의 확인, 평가영역과 내용선정, 평가방법(도구) 선정 및 개발, 평가시행 계획수립과 준비, 평가자료 수집과 시행, 평가결과 분석과 해석, 평가결과 보고와 활용, 메타 평가라고 하는 8단계로 나누고 있다.

3. 평가도구의 구비조건

교육은 바람직한 인간 행동 특성의 계획적인 변화이다(정범모, 1976). 평가도구는 평가활동에 있어서 없어서는 안 될 중요한 도구이지만 평가의 목적을 달성하기 위해서는 반드시 갖추어야 할 조건이 있다. 어떤 평가도구를 사용하느냐 하는 것은 평가의 결과에 심각한 영향을 미칠 수 있기 때문에 평가도구가 갖추어야 할 조건을 제대로 갖추고 있는지를 점검해야 한다. 평가도구가 구비해야 할 조건으로 타당도, 신뢰도, 객관도, 실용도를 들 수 있다.

가. 타당도

타당도란 평가도구가 측정하려고 하는 대상의 내용 그 자체를 재고 있는 충실성의 정도를 가리키는 말로 평가도구의 정확성과 관련된 문제이긴 하나 검사의 신뢰도처럼 재려고 하는 것이 아니라 무엇(what) 즉, 평가도구가 측정하고자 하는 것을 얼마나 충실히 측정하였는가를 나타내는 평가도구 목적의 적합성에 해당된다.

타당도를 이해하기 위해서는 다음과 같은 사항에 주목할 필요가 있다(성태제, 1996). 타당도는 피검사자 집단에 사용된 검사 도구에 의해서 얻어진 검사결과의 해석에 대한 적합성이지 검사자체와 관련된 것이 아니다. 그러므로 편의상 검사의 타당도라 표현하지만 정확히 말하여 검사결과로부터 만들어진 해석에 대한 타당성을 말하는 것이다. 또한 타당도는 정도의 문제이다. 타당도 근거에 따라 구체적으로 검사가 측정하고자 하는 특정한 목적에 비추어 '타당도가 높다' 또는 '타당도가 낮다'와 같이 평가하는 것이 정확한 표현이다. 타당도는 특별한 목적이나 해석에 제한된다. 한 검사가 모든 목적에 부합될 수 없듯이 이 검사는 무엇을 측정하는데 타당하다고 표현해야 한다.

나. 신뢰도

신뢰도란 측정하고자 하는 내용을 어느 정도 일관성 있게 측정하고 있는지를 알려주는 평가도구의 일관성 또는 정확성을 의미한다. 즉, 평가대상이 무엇이던 그것을 어느 정도 정확하게 측정하느냐에 관심을 가지고 평가과정에 작용하는 오차에 집중한다. 그러므로 동일한 평가도구를 동일한 대상에게 여러 번 반복해서 측정한 결과에서 각각 다른 수치가 나타날 때 오차의 변산이 커짐을 뜻하고 이는 곧 측정되어 나온 수치들의 신뢰는 물론 평가도구 자체의 신뢰도가 없음을 나타낸다(임진권, 1985). 신뢰도는 평가하고자 하는 것을

얼마나 안정적으로 일관성 있게 평가하느냐 하는 것이며 평가도구가 얼마나 정확하게 오차 없이 평가하느냐에 관한 문제이다. 따라서 신뢰도는 평가도구의 믿음성, 안정성, 일관성, 정확성 등을 나타내는 데 쓰인다.

타당도가 평가도구의 필수불가결한 조건이라 한다. 그러나 신뢰도가 낮으면 타당도는 이에 따라 낮아지기 때문에 신뢰도는 타당도의 선행조건이 된다. 즉, 신뢰도 없이 타당도 높은 평가도구는 존재할 수 없다. 타당도와 신뢰도가 같은 상태인 것이 가장 이상적인 평가도구이지만 그것은 거의 불가능하며 완벽한 측정이란 존재하기 어렵다.

다. 객관도

객관도란 채점자에 의해서 좌우되는 신뢰도의 일종으로 채점자 또는 검사자의 신뢰도라고 한다. 한 가지 반응 결과에 여러 사람의 채점 및 평가가 일치하는 같은 대상에 대해 다른 평가를 나타내는 것은 '평가자간 객관도', 한 평가자가 시간적 간격이나 생활의 차이에 따라 같은 대상에 대해 다른 평가결과를 나타내는 것은 '평가자내 객관도'라고 한다. 채점자의 채점이 어느 정도 신뢰를 주고 일관성 있느냐로 규정된다.

라. 실용도

실용도란 평가도구의 실용적 가치의 정도를 의미한다. 즉, 검사도구가 경비, 시간, 노력 등을 적게 들여서 평가의 목적을 충실하게 달성할 수 있어야만 실용성이 있는 것이다. 타당도, 신뢰도, 객관도가 높은 검사가 필요한 결정을 내리는 데 아무리 좋은 정보를 제공해 준다고 해도 그 검사를 실제로 이용하는 데 비용이나 노력이 많이 들면 활용하기가 어렵다. 따라서 실용도를 높이기 위해서는 검사의 실시와 채점이 쉬워야 하며 해석과 활용이 용이해야 하고 비용, 시간, 노력 등이 절약되어야 한다.

2. 내용 타당도의 개념과 검증방법

내용타당도란 한 검사를 구성하고 있는 문항들이 그 검사가 측정하고자 하는 내용을 측정하기 위해 만들어질 수 있는 문항 전집을 대표할 수 있도록 문항들이 표집 되어 있는 정도를 말한다. 즉, 검사의 문항들이 측정을 위하여 규정된 내용영역이나 또는 전체를 얼마나 잘 대표하느냐의 정도를 나타낸다. 내용타당도는 성취도 검사와 같이 측정하고자 하는 대상이 비교적 구체적으로 알려져 있을 때 채택할 수 있는 유용한 개념이다. 검사가 측정하고자 하는 내용을 골고루 측정할 수 있도록 문항을 얼마나 잘 표집 했느냐의 판단, 즉 내용타당도의 판단은 검사가 측정하고자 하는 분야의 전문가에 의해 이루어진다. 검사개발자의 주관적 판단에 의해 선택된 문항이 다른 전문가에 의해 역시 적합한 것으로 판단될 경우 객관적 절차에 의해 주어진 검사가 내용타당도의 증거를 가지고 있는 것으로 여길 수 있을 것이다. 따라서 학업성취도 평가에서 교과, 교수타당도를 증진시키기 위하여 내용과 행동영역으로 나누는 이원분류표의 작성이 매우 중요하다. 학업성취도 평가에 대한 평가내용에 기초한 근거의 검증은 문항이 검사 제작 전 작성한 평가 목표 이원분류표에 의하여 제작되었는지를 확인하는 방법을 사용한다. 전문가들의 판단에 의하여 평가의 타당성을 입증 받게 되므로 평가 목적의 적합성 여부를 검증할 수 있는 장점이 있다.

내용타당도를 객관적으로 분석하기 위한 두 가지 방법이 있다. 먼저 첫 번째 방법의 다

음과 같은 단계이다(황정규, 1973).

첫째, 내용의 전집을 명확하고 상세하게 정의하여 몇 개의 적합한 하위 영역으로 분류하고 각 영역의 중요성에 따라 적당한 비중을 각각 결정한다.

둘째, 검사 제작자는 각 문항에 대해서 그 문항이 재고 있는 내용과 기능을 명세화하고 그에 따라 문항의 상대적인 비중을 결정해야 한다.

셋째, 그 검사의 전반적인 타당도를 측정해 주는 기준을 만들고 또한 필요에 따라서는 평가하고자 하는 기능 및 중요한 내용 또는 내용에 대한 평가형식의 적합성을 측정하는 형식을 만들 수도 있을 것이다.

넷째, 그 검사문항의 내용타당도는 그 검사 문항을 각각 조사하고 각 문항에 대한 분류와 검사 제작자의 분류를 비교할 수 있을 것이다.

또 다른 방법은 수업목표와 평가목표를 Bloom(1956)이 제시한 내용-행동으로 이원 분류하여 내용타당도를 판단하는 것이다. 이를 단계별로 제시하면 다음과 같다.

첫째, 수업목표와 평가목표를 객관성 높게 내용-행동차원으로 분류할 교육전문가를 선정한다.

둘째, 수업목표를 내용과 행동차원으로 Bloom이 제시한 목표분류에 의거하여 이원분류표를 작성한다.

셋째, 평가문항을 내용과 행동차원으로 Bloom이 제시한 목표분류에 의거하여 이원분류표에 작성한다.

넷째, 수업목표의 분류와 검사문항의 분류와의 일치하는 정도가 크다면 이 검사는 내용타당도가 높다고 볼 수 있다.

그러므로 내용타당도는 주관적인 것이기는 하나 검사내용의 분석에 의하여 내용타당도가 '높다' 혹은 '낮다'로 표현함은 검사내용 전문가의 전문지식에 의존하는 것이기 때문에 임의적 판단이 아님을 전제한다.

III. 연구 방법

1. 연구대상

본 연구의 분석 대상은 초등학교 3-가 수학과 교사용 지도서에 제시된 1~6단원까지의 60개의 수업목표와 서울시 소재한 초등학교 중 지역별 다양성을 고려하여 선정한 5개 학교의 교사 제작 수학과 3-가 학업 성취도 평가도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가도구이다. 분석대상이 된 수업목표와 학업 성취도 평가 도구의 평가문항 수를 주제별로 분류한다.

2. 연구방법

지역별 다양성을 고려하여 선정한 서울시 소재한 5개의 초등학교 교사들이 제작한 수학과 3-가의 학업 성취도 평가 도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가 도구를 수집하였다. 연 2회 실시하는 단위 학교별 교사 제작 학업 성취도 평가 도구로 평가 범위는 1~6단원으로 실시한 것을 수집하였다. 수학과 3-가 교사용 지도서에 제시된 수업 목표 목록을 작성하고 5개교의 교사 제작 학업 성취도

평가 도구의 평가 목표를 수집하였다. 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가 도구의 평가 목표는 수학과 3-가 1~6단원에 해당하는 평가 문항을 선별하고 2007년 초등학교 3학년 국가수준 기초학력 진단평가 연구 보고서를 바탕으로 작성하였다. 수학과 3-가 1~6단원의 내용과 상이한 평가 문항은 목록에서 생략했다.

객관성 높은 이원분류표를 작성하기 위하여 교육과정을 전공하고 수학과 3-가 학업 성취도 평가 도구의 제작 경험이 있는 초등학교 교사 4인과 본 연구자 즉, 5인을 목표분류자로 선정했다. 2008년 7월에서 2월까지 본 연구자와 분류자 개별적인 이원분류표 수집을 위한 협의 1회와 각자 다른 의견을 협의하기 위하여 3번에 걸친 협의 과정 즉, 총 4번에 걸쳐 본 분류자 5인은 이원분류표를 함께 작성하였다. 3-가 교사용지도서에 제시된 수업 목표를 5인의 분류자들의 합의에 의하여 수업목표·평가목표 이원분류표를 작성하였다. 수업 목표와 평가목표를 이원분류하는 과정에서 4인 이상 합의한 것은 ◎로 나타내었고 4인 미만 합의는 다수의 합의자가 동의한 의견으로 합의하였으며 ○로 나타내었다.

<표 1> 수학과 3-가 수업목표 이원분류표(1~6단원) 예시

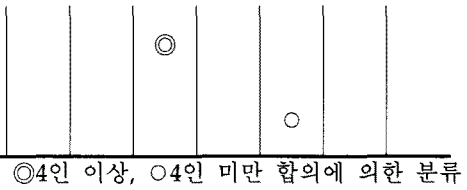
내용	행동	지식	이해	적용	분석	종합	평가	계
1. 천의 크기 알아보기			○					13
2. 천을 읽고 쓰기	◎							
3. 1000개씩 묶음의 수를 세기			◎					
4. 몇 천을 읽고 쓰기	◎							
5. 네 자리 수의 기수법의 원리 이해하기			○					
6. 네 자리 수를 읽고 쓰기	◎							
7. 네 자리 수에서 수의 값, 자리의 숫자 알기			○					
8. 수세기를 통하여 네 자리 수의 계열 알기					◎			
9. 뛰어 세기를 통하여 계열 알기					◎			
10. 9999 다음 수로서 10000을 알기					○			
11. 네 자리 수의 크기 비교하기				○				
12. 생활에서 네 자리 수의 예 찾아보기						○		
13. 놀이(숫자카드)를 통하여 네 자리 수의 크기 비교하기				◎				

◎4인 이상, ○4인 미만 합의에 의한 분류

<표 2> 학업 성취도 평가도구 평가목표 이원분류표(1~6단원) 예시

내용	행동	지식	이해	적용	분석	종합	평가	계
1. 네 자리 수 읽고 쓰기		◎						
2. 몇 천을 읽고 쓰기		◎						
3. 놀이(숫자카드)를 통하여 네 자리 수의 크기 비교하기					◎			
4. 네 자리 수의 크기 비교하기					○			
5. 받아올림이 있는 세 자리 수 덧셈 계산하기			◎					
6. 받아내림이 있는 세 자리 수 뺄셈 계산하기			◎					

- 7. 받아내림이 있는 세 자리 수 뿔셈 생활장면에서 해결하기
- 8. 여러 가지 방법으로 받아내림이 있는 세 자리 수 차 구하기



2. 연구절차

내용 타당도를 분석하고 비교하기 위하여 수업목표 이원분류표에 의거하여 수업목표의 행동차원별 빈도수와 백분율(%), 내용차원별 빈도수와 백분율(%)을 구하였다. 그리고 평가목표의 행동차원별 빈도수와 백분율(%) 및 내용차원별 빈도수와 백분율(%)을 구하였다. 마지막으로 학업 성취도 평가 도구의 내용 타당도 검증을 위해 수업목표와 평가목표의 행동차원에서의 수준별 비율차와 내용차원에서 주제별 비율차를 검증하였다.

내용 타당도 분석을 위한 비율차 검증은 다음과 같은 방법으로 계산한다.

$$t = Dp / \sigma Dp$$

t : 비율차 지수

Dp : 비율차

σDp : 비율의 표준오차

$$\sigma Dp = \sqrt{pq \left(\frac{1}{M} + \frac{1}{N} \right)}$$

$$p = \frac{M P_1 + N P_2}{M + N}$$

$$q = (1-p)$$

N1 = 수업목표 수

N2 = 평가목표 수

P1 = 수업목표 백분율

P2 = 평가목표 백분율

통계처리 결과 비율차의 유의도가 $p > .05$ 이면 수업목표가 평가목표와 동질적이며 내용 타당도가 높은 것으로 해석하였다. 평가 도구의 내용 타당도는 <표 3>과 같이 내용 타당도 판정 기준에 의해 판단하였다.

<표 3> 내용 타당도 판정 기준

유의도 $p < .05$ 인 영역의 수	내용 타당도
0~2개	높다
3~4개	보통
5개 이상	낮다

5개교의 학업 성취도 평가 도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가 도구의 내용 타당도를 각각 분석해 보았다. 5개교의 교사 제작 학업 성취도 평가 도구의 내용 타당도를 종합하여 교사 제작 학업 성취도 평가 도구와 한국교육과정평가원(KICE)의 내용 타당도를 비교해 보았다.

IV. 연구 결과

1. 수업목표 분석

3-가 수학과 1~6단원까지의 수업목표를 이원 분류한 결과 행동 영역에서 지식수준이 16.7%, 이해력이 35%, 적용력이 15%, 분석력이 15%, 종합력이 16.7%, 평가력이 1.6%이었다. 지식과 이해의 목표는 전체의 51.7%, 고차적 지적능력을 평가하는 적용력, 분석력, 종합력, 평가력의 목표는 전체의 48.3%로 분석되었다.

내용 영역에서는 10000까지의 수가 21.7%, 덧셈과 뺄셈 21.7%, 평면도형 20%, 나눗셈 9%, 도형 움직이기 8.3%, 곱셈 13.3%였다.

2. 평가목표 분석

한국교육과정평가원(KICE) 제작 평가 도구의 평가 목표를 이원 분류한 결과 행동 영역에서 지식수준이 17.7%, 이해력이 52.9%, 적용력이 29.4%, 분석력, 종합력, 평가력은 평가 문항이 제시되지 않았다. 지식과 이해의 목표는 전체는 70.6%, 고차적 지적능력을 평가하는 적용력, 분석력, 종합력, 평가력의 목표는 전체의 29.4%로 분석되었다.

내용 영역에서는 10000까지의 수가 11.8%, 덧셈과 뺄셈 23.4%, 평면도형 29.4%, 나눗셈 11.8%, 도형 움직이기 11.8%, 곱셈 11.8%였다.

교사 제작 학업 성취도 평가 도구의 평가 목표를 이원 분류한 결과 행동 영역에서 지식수준이 15.1%, 이해력이 40.3%, 적용력이 20.2%, 분석력 11.8%, 종합력 10.9%, 평가력 1.7%로 분석되었다. 지식과 이해의 목표는 전체는 55.4%, 고차적 지적능력을 평가하는 적용력, 분석력, 종합력, 평가력의 목표는 전체의 44.6%로 분석되었다.

내용 영역에서는 10000까지의 수가 15.1%, 덧셈과 뺄셈 17.7%, 평면도형 16%, 나눗셈 16.8%, 도형 움직이기 15.1%, 곱셈 19.3%였다.

3. 행동차원에서 내용 타당도 분석

행동차원에서 학업 성취도 평가 도구의 평가도구별 내용 타당도 분석은 수업목표와 평가목표의 행동영역별 비율을 비율차 검증하였다. 한국교육과정평가원(KICE) 제작 평가 도구와 교사 제작 평가 도구(A~E)의 내용 타당도를 분석한 결과는 다음과 같다.

가. 한국교육과정평가원(KICE) 제작

한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구를 행동차원에서의 내용 타당도를 분석한 결과는 <표 4>와 같다.

<표 4> 행동영역의 수업목표와 평가목표의 비율차 검증
- 한국교육과정평가원(KICE) 제작

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	17.7	0.010	0.247	0.040
이해력	35	52.9	0.179	0.197	0.909
적용력	15	29.4	0.144	0.224	0.643
분석력	15	.	0.150	0.119	1.261
종합력	16.7	.	0.167	0.118	1.415
평가력	1.6	.	0.016	0.125	0.128

* p < .05

행동차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 한국교육과정평가원(KICE) 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

나. 교사 제작

(1) A 교

A 교 교사 제작 학업 성취도 평가 도구를 행동차원에서 내용 타당도를 분석한 결과는 <표 5>와 같다.

<표 5> 행동영역의 수업목표와 평가목표의 비율차 검증: A 교

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	20	0.033	0.210	0.157
이해력	35	40	0.005	0.190	0.026
적용력	15	24	0.090	0.306	0.294
분석력	15	4	0.110	0.692	0.159
종합력	16.7	12	0.047	1.278	0.037
평가력	1.6	.	0.016	0.788	0.020

* p < .05

행동차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 A 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(2) B 교

B 교 교사 제작 학업 성취도 평가 도구를 행동차원에서 내용 타당도를 분석한 결과는 <표 6>과 같다.

<표 6> 행동영역의 수업목표와 평가목표의 비율차 검증: B 교

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	24	0.073	0.204	0.358
이해력	35	40	0.050	0.181	0.276
적용력	15	8	0.070	0.269	0.260
분석력	15	16	0.010	0.216	0.046
종합력	16.7	8	0.087	0.279	0.312
평가력	1.6	4	0.024	0.233	0.103

* $p < .05$

행동차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 B 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(3) C 교

C 교 교사 제작 학업 성취도 평가 도구를 행동차원에서 내용 타당도를 분석한 결과는 <표 7>과 같다.

<표 7> 행동영역의 수업목표와 평가목표의 비율차 검증: C 교

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	12.5	0.042	0.239	0.178
이해력	35	50	0.150	0.179	0.838
적용력	15	25	0.100	0.207	0.483
분석력	15	8.3	0.067	0.270	0.248
종합력	16.7	.	0.167	0.118	1.415
평가력	1.6	4.2	0.026	0.237	0.110

* $p < .05$

행동차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 C 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(4) D 교

D 교 교사 제작 학업 성취도 평가 도구를 행동차원에서 내용 타당도를 분석한 결과는 <표 8>과 같다.

<표 8> 행동영역의 수업목표와 평가목표의 비율차 검증: D 교

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	12	0.047	0.239	0.197
이해력	35	32	0.030	0.198	0.152
적용력	15	24	0.090	0.205	0.439
분석력	15	16	0.110	0.216	0.509
종합력	16.7	16	0.007	0.220	0.032
평가력	1.6	.	0.016	0.125	0.128

* p < .05

행동차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 D 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(5) E 교

E 교 교사 제작 학업 성취도 평가 도구를 행동차원에서 내용 타당도를 분석한 결과는 <표 9>와 같다.

<표 9> 행동영역의 수업목표와 평가목표의 비율차 검증: E 교

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	5	0.117	0.381	0.307
이해력	35	40	0.050	0.201	0.249
적용력	15	20	0.050	0.223	0.224
분석력	15	15	0	0.238	0
종합력	16.7	20	0.033	0.225	0.147
평가력	1.6	.	0.016	0.125	0.128

* p < .05

E교 교사 제작 학업 성취도 평가도구를 행동차원에서 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 E 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(6) 종합

교사 제작 학업 성취도 평가 도구(A~E)를 종합하여 행동영역에서 내용 타당도를 분석한 결과는 <표 10>과 같다.

<표 10> 행동영역의 수업목표와 평가목표의 비율차 검증: 종합

구 분	수업목표	평가목표	Dp	σDp	t
지 식	16.7	15.1	0.016	0.143	0.112
이해력	35	40.3	0.053	0.130	0.408
적용력	15	20.2	0.052	0.153	0.340
분석력	15	11.8	0.032	0.144	0.222
종합력	16.7	10.9	0.058	0.143	0.406
평가력	1.6	1.7	0.001	0.158	0.006

* p < .05

<표 10>과 같이 교사 제작 학업 성취도 평가도구를 행동차원에서 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 B 교 교사 제작 학업 성취도 평가 도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

2. 내용차원에서 내용 타당도 분석

내용차원에서 학업 성취도 평가 도구의 평가도구별 내용 타당도 분석은 수업목표와 평가목표의 내용영역별 비율을 비율차 검증하였다. 한국교육과정평가원(KICE) 제작 평가 도구와 교사 제작 평가 도구(A~E)의 내용 타당도를 분석한 결과는 다음과 같다.

가. 한국교육과정평가원(KICE) 제작

한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구를 내용차원에서의 내용 타당도를 분석한 결과는 <표 11>과 같다.

<표 11> 내용영역의 수업목표와 평가목표의 비율차 검증
: 한국교육과정평가원(KICE) 제작

구 분	수업목표	평가목표	Dp	σDp	t
10000까지의 수	21.7	11.8	0.099	0.306	0.324
덧셈과 뺄셈	21.7	23.4	0.017	0.233	0.073
평면도형	20	29.4	0.094	0.223	0.422
나눗셈	15	11.8	0.032	0.275	0.116
도형 움직이기	8.3	11.8	0.035	0.243	0.144
곱셈	13.3	11.8	0.015	0.266	0.056

* p < .05

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 한국교육과정평가원(KICE) 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정

기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

나. 교사 제작

(1) A 교

A 교 교사 제작 학업 성취도 평가 도구를 내용차원에서 내용 타당도를 분석한 결과는 <표 12>와 같다.

<표 12> 내용영역의 수업목표와 평가목표의 비율차 검증: A 교

구 분	수업목표	평가목표	Dp	oDp	t
10000까지의 수	21.7	16	0.057	0.230	0.248
덧셈과 뺄셈	21.7	16	0.057	0.230	0.248
평면도형	20	16	0.040	0.226	0.177
나눗셈	15	16	0.010	0.216	0.046
도형 움직이기	8.3	16	0.077	0.216	0.356
곱셈	13.3	20	0.067	0.208	0.322

* p < .05

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 A 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(2) B 교

B 교 교사 제작 학업 성취도 평가 도구를 내용차원에서 내용 타당도를 분석한 결과는 <표 13>과 같다.

<표 13> 내용영역의 수업목표와 평가목표의 비율차 검증: B 교

구 분	수업목표	평가목표	Dp	oDp	t
10000까지의 수	21.7	12	0.097	0.248	0.391
덧셈과 뺄셈	21.7	12	0.097	0.248	0.391
평면도형	20	20	0	0.192	0
나눗셈	15	16	0.01	0.117	0.085
도형 움직이기	8.3	16	0.077	0.164	0.470
곱셈	13.3	24	0.107	0.157	0.682

* p < .05

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 B 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하

여 내용 타당도가 매우 높다고 해석하였다.

(3) C 교

C 교 교사 제작 학업 성취도 평가 도구를 내용차원에서 내용 타당도를 분석한 결과는 <표 14>와 같다.

<표 14> 내용영역의 수업목표와 평가목표의 비율차 검증: C 교

구 분	수업목표	평가목표	Dp	σ Dp	t
10000까지의 수	21.7	16.66	0.0504	0.231	0.218
덧셈과 뺄셈	21.7	16.66	0.0504	0.231	0.218
평면도형	20	16.66	0.0334	0.227	0.147
나눗셈	15	16.66	0.0166	0.218	0.076
도형 움직이기	8.3	16.66	0.0836	0.218	0.383
곱셈	13.3	16.66	0.0336	0.215	0.156

* $p < .05$

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 C 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(4) D 교

D 교 교사 제작 학업 성취도 평가 도구를 내용차원에서 내용 타당도를 분석한 결과는 <표 15>와 같다.

<표 15> 내용영역의 수업목표와 평가목표의 비율차 검증: D 교

구 분	수업목표	평가목표	Dp	σ Dp	t
10000까지의 수	21.7	16	0.057	0.217	0.263
덧셈과 뺄셈	21.7	24	0.023	0.181	0.127
평면도형	20	12	0.080	0.129	0.620
나눗셈	15	24	0.090	0.162	0.556
도형 움직이기	8.3	8	0.003	0.214	0.014
곱셈	13.3	16	0.027	0.185	0.146

* $p < .05$

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 유의 있는 차이가 없었다. 따라서 D 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(5) E 교

E 교 교사 제작 학업 성취도 평가 도구를 내용차원에서 내용 타당도를 분석한 결과는 <표 16>과 같다.

<표 16> 내용영역의 수업목표와 평가목표의 비율차 검증: E 교

구 분	수업목표	평가목표	Dp	σDp	t
10000까지의 수	21.7	15	0.067	0.258	0.260
덧셈과 뺄셈	21.7	15	0.067	0.258	0.260
평면도형	20	15	0.050	0.253	0.198
나눗셈	15	10	0.050	0.272	0.184
도형 움직이기	8.3	20	0.117	0.229	0.511
곱셈	13.3	25	0.117	0.218	0.537

* p < .05

내용차원에서의 내용타당도를 분석한 결과 모든 영역에서 의의 있는 차이가 없었다. 따라서 E 교 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

(6) 종합

교사 제작 학업 성취도 평가 도구(A~E)를 종합하여 내용영역에서 내용 타당도를 분석한 결과는 <표 17>과 같다.

<표 17> 내용영역의 수업목표와 평가목표의 비율차 검증: 종합

구 분	수업목표	평가목표	Dp	σDp	t
10000까지의 수	21.7	15.1	0.066	0.136	0.485
덧셈과 뺄셈	21.7	17.7	0.040	0.139	0.288
평면도형	20	16	0.040	0.136	0.294
나눗셈	15	16.8	0.018	0.148	0.122
도형 움직이기	8.3	15.1	0.067	0.173	0.387
곱셈	13.3	19.3	0.060	0.157	0.382

* p < .05

<표 17>과 같이 교사 제작 학업 성취도 평가도구를 행동차원에서 내용타당도를 분석한 결과 모든 영역에서 의의 있는 차이가 없었다. 따라서 교사 제작 학업 성취도 평가도구는 <표 3>의 내용타당도 판정 기준에 의거하여 내용 타당도가 매우 높다고 해석하였다.

한국교육과정평가원(KICE)와 교사가 제작한 학업 성취도 평가 도구의 행동·내용차원에

서 내용 타당도의 유의미한 차이를 분석한 결과 행동차원에서 교사 제작 수학과 3-가 학업 성취도 평가도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가도구의 내용 타당도를 분석·비교한 결과, 6개 영역(지식, 이해력, 적용력, 분석력, 종합력, 평가력)에 의의 있는 차이가 없었다($p < .05$). 그러므로 두 제작 모두 내용 타당도가 매우 높은 것으로 나타났다.

내용차원에서 교사 제작 수학과 3-가 학업 성취도 평가도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가도구의 내용 타당도를 분석·비교한 결과, 6개 영역(지식, 이해력, 적용력, 분석력, 종합력, 평가력)에 의의 있는 차이가 없었다($p < .05$). 그러므로 두 제작 모두 내용 타당도가 매우 높은 것으로 나타났다.

그러므로 교사 제작 수학과 3-가 학업 성취도 평가도구와 한국교육과정평가원(KICE)에서 제작한 3학년 국가수준 기초진단 학업 성취도 평가도구는 행동차원과 내용차원에서 모두 내용타당도가 높은 것으로 결론지을 수 있다.

V. 결 론

본 연구의 목적은 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구와 교사 제작 학업 성취도 평가 도구의 내용타당도를 분석·비교하는데 있으며, 학교 현장에서 실시되는 학업 성취도 평가 도구의 개선을 위한 기초 자료를 제시하는데 그 의의가 있다.

연구의 목적을 달성하기 위하여 행동차원에서 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구와 교사 제작 학업 성취도 평가 도구의 내용타당도를 분석한 결과 모든 수집된 학업 성취도 평가 도구 모두 지식, 이해력, 적용력, 분석력, 종합력의 모든 영역에서 의의 있는 차이가 없었으므로 내용타당도가 높다고 결론지을 수 있다. 그러나 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구는 기초학력 진단평가로 기초학력의 성격을 고려하여 구성된 점을 고려하여도 행동 영역은 계산, 이해, 추론 및 문제 해결 중심으로 구성되어 지식, 이해력을 평가하는데 평가 문항이 집중되어 있고 고차적 지적 능력인 적용력을 제외한 분석력, 종합력, 평가력은 평가문항이 제시되지 않아서 편중된 평가 문항의 출제 경향을 나타내었다. 교사 제작 학업 성취도 평가 도구는 수업목표와 평가목표의 분포가 크게 차이가 나지 않았으나 각 학교별 다소 편차가 있는 것으로 보여 지고 있다. 평가력은 Bloom이 제시한 목표 분류 중 가장 상위 수준으로 초등학교 평가에서 다루어지기 어렵지만 창의력 문제 해결력을 추구하기 위해 포함되는 것이 바람직하겠다. 또한 평가력은 평가 문항이 제시되지 않은 평가 도구도 상당 수 있어 출제자의 주의가 요구된다.

내용차원에서 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구와 교사 제작 학업 성취도 평가 도구의 내용타당도를 분석한 결과 모든 수집된 학업 성취도 평가 도구 모두 지식, 이해력, 적용력, 분석력, 종합력의 모든 영역에서 의의 있는 차이가 없었으므로 내용 타당도가 높다고 결론지을 수 있다. 한국교육과정평가원(KICE) 제작 학업 성취도 평가 도구는 측정, 도형, 수와 연산으로 정하여 분류하면 비슷한 분포를 나타내지만 덧셈과 뺄셈, 평면 도형에 다른 내용보다 비중 있게 평가 문항을 출제하였다. 이는 30문항 중 3-가의 1~6단원 이외에 다른 단원도 포함되어 이와 같이 나타난 것을 생각된다. 교사 제작 학업 성취도 평가 도구는 내용 주제별로 고르게 평가 문항이 제시되었다. 학업 성취도 평가를 특정 영역에 치우침 없이 고른 비율로 제작하는 것은 바람직한 평가 도구이다. 그러나 내

용 주제별로 제시된 평가 문항이 대부분 단원을 기준으로 순차적으로 제시된 경향을 보여 학생이 학업성취도 평가 문항을 예측할 가능성이 있다. 따라서 이를 좀 더 다양한 방법으로 제시하도록 개선할 필요가 있다.

본 연구는 수학과 3-가에 제한하여 검증하였으나, 전 교과 영역으로 확대하여 학교 현장에서 이루어지고 있는 평가 도구의 내용타당도를 분석하여 전 교과 전 영역의 학업성취도 평가가 수업목표에 맞게 평가되고 있는지 검증해 볼 필요가 있다.

교사 제작 학업 성취도 평가도구는 내용타당도는 높으나 수업 목표의 배열과 유사하게 평가 문항이 배열되어 있다. 이는 학생이 평가 문항에 대하여 예측하여 문제 해결을 할 수 있도록 하므로 보다 세심한 성취도 평가도구의 제작이 필요하다. 또한 평가 문항이 지식, 이해력, 적용력을 평가하는데 집중된 경향을 보여 창의력 신장을 위한 고등정신기능을 기르는 평가가 이루어지도록 평가 방법 개선 방안과 평가 도구 개발에 힘써야 할 것이며 각 학교간의 편차를 해소하기 위하여 많은 교사들의 상호 교류도 활발히 진행되어야 할 것임을 제언한다.

참 고 문 헌

- 강기훈 (2005). 통계학 개론-엑셀을 이용한 실습. 서울: 자유아카데미.
- 강승호, 김명숙, 김정환, 남현호 (1996). 현대 교육 평가의 이론과 실제. 서울: 양서원.
- 강완 외 (1999). 초등 수학 학습지도의 이해. 서울: 양서원.
- 강현석, 주동범 (2004). 현대 교육과정과 교육평가. 서울: 학지사.
- 교육 인적 자원부 (2001). 수학 3-가 초등학교 교사용 지도서. 서울: 대한교과서 주식회사.
- 구정선 (2007). 사회과 학업성취도 문항 분석. 한국교원대 교육대학원 석사학위논문.
- 김대현, 김석우 (2005). 교육과정 및 교육평가. 서울: 학지사.
- 김민경 외 (2006). 초등교재연구. 서울: 교육과학사.
- 김정숙 (2005). 교사 제작 평가 문항의 내용타당도 분석. 한국교원대 교육대학원 석사학위 논문.
- 김주아 (2001). 학교 평가에서의 질적 접근 방안 연구. 이화여자대학교 교육대학원 석사학위 논문.
- 김진규 (2002). 교육과정과 교육평가. 서울: 동문사.
- 박도순 (2007). 교육평가-이해와 적용. 서울: 교육과학사.
- 박도순, 홍후조 (1987). 교육과정과 교육평가. 서울: 문음사.
- 박수만 (2007). Bloom의 인지적 영역에 기반한 대학수학능력시험 수리영역 문항분석. 인제대학교 교육대학원 석사학위 논문.
- 변창진, 최진승, 문수백, 김진규, 권대훈 (1996). 교육평가. 서울: 학지사.
- 성태제 (2002). 타당도와 신뢰도. 서울: 학지사.
- 성태제 (2002). 현대교육평가. 서울: 학지사.
- 송준호 (2003). 교사제작 총괄평가의 내용타당도 분석. 한국교원대 교육대학원 석사학위 논문.
- 심수정 (2006). 학업 성취도 평가 도구의 제작 주체별 내용 타당도 비교. 한국교원대 교육대학원 석사학위 논문.
- 이칭찬, 황선경 (2003). 교육과정과 교육평가. 서울: 동문사.
- 이해명 외 (2006). 현대 교육과정 및 평가. 서울: 교육아카데미.
- 임진권 (1985). 교육평가. 서울: 대은출판사.
- 정범모 (1976). 교육평가의 원리. 서울: 풍국학원.
- 정종진 (2002). 교육평가의 원리. 서울: 교육출판사.
- 정종진 (1999). 교육평가의 이해. 서울: 양서원.

- 조승제 (2007). *교육과정과 평가의 쟁점*. 서울: 교육과학사.
- 한국교육과정평가원 (2005). *교사의 학생평가 전문성 신장 연구(II)자료집: 교사의 학생평가 전문성 기준*. 한국교육과정평가원.
- 한국교육과정평가원 (2008). *2007년 초등학교 3학년 국가수준 기초학력 진단평가 연구-기초수학-*. 한국교육과정평가원.
- Bloom, B. S., Engelhart, M. D., Frust, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of education objective, Hand Book I ; Cognitive Domain*. NY; Mckey.
- 임의도 외(역). *교육목표 분류학*. 서울: 교육과학사(1984).
- Tyler. R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: University of Chicago Press.

<Abstract>

An Analysis of Content Validity of Third-Grade Mathematics Achievement Tests

Kim, Eunah³); & Kang, Wan⁴)

The purpose of this study is to find out whether Achievement Tests are fully performing their role as an objective standard that measures student's educational achievement level by analysing the content validity of Achievement Tests developed by KICE and teachers at elementary school.

In the study, the content validity of achievement tests were analyzed in the behavioral content objective dimensions.

60 instructional objectives from the Unit one to the Unit six contained in the teachers' guidebook for the elementary third-grade Math subject were analyzed into dimensions of behavior and content. And the Achievement Test developed by KICE and teachers in five elementary schools randomly chosen were collected and analyzed. Then, differences of the proportion between instructional objectives and evaluative objectives in each dimension statistically were verified.

The results of the study are as follows :

1. In the dimension of behavior, as analysing and comparing the content validity of achievement tests, there was no significant difference in all domains(knowledge, comprehension, application, analysis, synthesis and evaluation). ($p < .05$) Therefore, it could be concluded that content validity of the Achievement Test is very high.

2. In the dimension of content, similarly there is no significant difference in all domains between achievement tests by both KICE and teachers. ($p < .05$) Therefore, the content validity of all tests are very high.

In conclusion, it could be concluded that content validity of achievement tests is considerably high in content and behavior dimension.

The study suggest the followings :

1. By expanding to the other subjects, there are needs to analyze and verify the content validity of achievement tests.

2. Even the content validity of achievement tests is considerably high however, achievement items are focused on evaluation 3 domains(knowledge, comprehension, application). Therefore evaluation evenly among 6 cognitive domains is required. And further to reduce the deviation of schools, there are needs to active interchange between teachers.

Keywords: evaluation, Achievement Tests, validity, content validity

논문접수: 2010. 04. 25

논문심사: 2010. 05. 12

게재확정: 2010. 08. 04

3) bubbbs@hanmail.net

4) wkang@snue.ac.kr