

다음은 지난 10월 26일~27일 양일간 본회 강당에서 전국 52개 간호교육기관 교수 54명이 참가한 가운데 개최된 간호원 국시문제 출제개선을 위한 워샷의 주제를 발췌 수록하였다. -편집자주-

## 교육평가의 원리

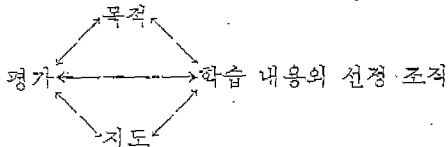
김            란            수    박사

<연세·문과대학 교수>

간호원 국가시험, 출제와 관련지어 교육평가의 원리와 지식 태도 기술평가의 원리에 대하여 이야기하겠다.

평가란 것은 가치판단이나 무엇을 판단하느냐 교육목적을 어느 정도 달성했느냐 하는 것에 대한 측정을 해 보고 그 결과에 대한 가치판단을 내리는 것이 평가라고 이야기할 수 있다. 즉 평가란 상당히 광범위한 뜻을 지닌 용어이다. 이것은 비단 학생의 입장에서 얼마만큼 우리가 교육하려고 했던 목적을 달성했느냐 하는 것을 평가라 하겠지만 우리가 교육하려고 했던 계획 자체가 얼마만큼 효율적이었느냐 하는 교육계획에 대한 평가도 포함하는 이야기로 받아들여져야 한다.

본래 교육의 process를 크게 4분류로 나누어 보면 다음의 그림을 참고할 수 있다.



위의 그림에서 제대로 우리가 process를 생각 해본다면 평가가 시발점이라 생각하면 좋을 것이다. 우리가 어떤 교육목표를 세울 것이냐 하는 것이 간호교육의 실재를 보고 또 간호원들이 나가서 일할 전문적인 활동에 대한 것을 사전평가 해서 이런 교육목표를 세워야 하겠다. 동시에 평가란 이 목적설정을 한 것이 위에서 말한대로 설정한 교육목적을 얼마만큼 달성했느냐 하는

것이다.

그러므로 목적설정도 즉각적인 영향을 주며 이것은 모든 과정이 이런 식으로 이루어진다고 생각해야 한다. 그중 어느 것이 먼저든 출발이 되느냐 하는 것은 보는 관점에 따라서 다르겠는데 우리가 제대로 교육목적을 세우려고 하면 일단은 간호교육의 실재를 평가해 보고 간호원들의 전문적인 업무수행에 대해 평가해 보고 사회가 무엇을 요구하느냐 등등을 모두 감안해서 교육목적을 세우고 계획을 세워야 하기 때문에 평가는 마지막 단계에 오는 것이 아니고 출발이라고 이야기할 수 있다. 사실 평가방법 자체도 평가를 해야 하고 평가 자체의 자체평가도 해야 한다. 그렇게 보면 모두가 평가가 아니냐 그렇게 이야기할 수도 있다. 우리가 교육을 하는 process는 뚜렷한 목적을 가지고 그것을 어떻게 하면 잘 달성할 수 있느냐 하는 의식때문에 언젠가지 자기성찰, 자기반성, 자기평가에 의식적인 노력이 뒤따르지 않을 때는 제대로 될 수 없다.

평가에 있어 제일 중요한 것은 목적설정이다. 이 목적 달성을 위해 어떤 교육경험을 줄 것이냐 또 이것을 어떤 방식으로 평가한 것이냐는 그 목적에 달려 있다. 그러므로 사실은 교육평가의 시발점은 교육목적 설정을 합리적으로 세우는 것이라고 말할 수 있다.

교육목적을 몇가지 차원으로 생각하면 다음과 같다.

이념

2. 사회현실
3. 간호학문의 추세
4. 평가
5. 학습 심리학적인 고찰

투철한 교육목적을 갖지 않고는 평가라는 것은 의미가 없다고 할 수 있고 또 그렇게 해서 교육이 제대로 될 수 없다. 평가는 사실상 이념적인 차원, 철학적인 측면까지도 포함을 해야 한다. 평가 그 자체가 Valuing 가치판단하는 것이기 때문에 상당히 많은 면에 신경을 써야 한다. 너무 기술적인 평가만 하나까 평가가 고식적이고 배너리즘에 빠지게 된다.

평가는 몇가지 기능을 갖고 있는데 첫째는 학습의욕을 고취시키는 기능을 갖고 있다. 그것이 긍정적이든 부정적이든 간에 평가는 그러한 기능을 발휘하고 있다. 그렇기 때문에 간호원 면허를 국가시험제도로 부여하는 것이 좋다, 나쁘다 등의 여러가지 얘기가 있을 수 있을 것이다. 그것은 왜냐하면 간호전문직에 있어서 지식의 교육이 얼마나 중요하여 이런 시험을 치느냐 혹은 이와 반대의 의견을 가진 사람들도 있겠지만 그러한 시험이 있음으로 인하여 전국의 간호교육 수준을 향상시키는데 많은 영향을 주었을 것이다.

둘째로 교육의 방향 제시적인 기능을 갖고 있다. 평가의 내용이 곧 교육의 내용이다. 즉 쉽게 말하면 국가시험에서 출제되지 않으면 가르칠 필요가 없다는 등의 교육방향이 될 수 있다는 것이다. 또한 평가방식이 교수 학습의 방식을 결정하는 것이 된다. 무엇을 어떻게 출제하느냐 하는 것은 교육에 미치는 영향이 크다. 바로 여러분들이 출제하는 그것이 바로 우리나라의 간호교육의 방향을 제시하는 것이 된다. 이렇게 생각하면 출제자의 책임은 막중한 것이다.

평가의 기능은 그밖에도 진단적인 기능이랄든가 선발적인 기능 등이 있지만 후에 언급하기로 하고 평가의 중요성을 평가의 기능선에서 언급해 보기로 한다.

평가는 비단 필답시험만으로 되는 것이 아니고 관찰을 통해서 등의 여러가지가 있지만 그중에서도 일반적으로 능력검사의 경우에는 제일 크게 대두되는 것이 필답고사라고 볼 수 있다.

평가를 하기 위해서는 측정을 해야 하며 측정을 하기 위해서는 검사도구가 있어야 한다. Test는 measurement를 위한 것이고 measurement

는 그 결과를 평가하기 위해 있는 것이다. 개념으로 보면 평가가 가장 큰 개념이고 그 다음이 측정이며 그 다음이 검사다. 그러나 실제적으로 우리는 필답고사는 워튼검사를 통해서 많이 하게 된다.

검사를 분류하면 상대적인 평가와 절대적인 평가로 분류할 수 있다. 다시 말하면 선발목적의 시험과 자격부여 목적의 시험으로 구분할 수 있다. 상대적인 평가의 입장은 100명이 시험을 치면 100명의 우선순위를 결정하기 위해 사용되며 우열의 순위검사라고 할 수 있으며 절대적인 평가는 일정한 기준(Criteria)을 놓고 거기에 도달여부를 평가하는 시험이다. 그러므로 100명 가운데 다 합격할 수도 다 합격되지 않을 수도 있는 입장에 있다.

상대적인 평가의 문제를 만들 때는 시험이 100명의 우열을 갈라 내는 목적이므로 시험문제가 쉬운 문제에서 어려운 문제로 고루 펼쳐 있어야 한다. 그러나 절대평가시험에 있어서는 적어도 간호교육을 받았으면 이 수준은 이해하고 있어야 한다는 일정선을 놓고 이것의 이해력을 따지는 것이다. 그러므로 간호원 국가시험은 절대적인 평가의 입장이 되어야 한다. 시험문제를 쉬운 문제에서 어려운 문제로 골고루 내어서는 안된다. 적어도 간호원이 된다고 할 때에는 이 개념은 이만한 깊이를 제대로 이해하고 있어야 된다는 판단이 내려야 한다. 그 기준은 출제자 속에 있을 수밖에 없다. 그렇기 때문에 간호원 국가시험출제자는 간호학의 전문가로서 권위가 인정될 수 있는 사람에게 제한해야 한다. 누구나 출제한다는 것은 안된다.

또 다른 방식으로 분류하면 속도검사(Speed test)와 power test로 분류할 수 있다. 대체로 norm referenced evaluation에 있어서 순위를 정할 때는 시험을 치루는데 대충일 때는 일일이 할 수가 없다. 간호원 국가시험에 있어서 원칙으로 자격고사의 성격은 power test 성격이어야 한다. 즉 문제를 많이 내서 얼마큼 많이 풀었느냐가 중요한 것이 아니고 기본적인 개념을 깊이 있게 이해했느냐 못했느냐를 알기 위해 문제수를 줄이고 논문식으로 해서 분체해결 역량이 있느냐 없느냐를 결정하는 것이며 time factor는 그렇게 중요한 것이 아니다.

응시자들이 절반밖에 풀 수 없는 시간을 준다. 면 guessing factor가 문제 된다. 아무렇게나

추측해서 답하는 경우가 많게 된다. 간호원 국가시험은 약 3,000명을 대상으로 하기 때문에 부득이 시간제한을 두어야 하는데 어느 시간이 적정성이나 응시자의 7~8할이 풀 수 있는 시점으로 출제분량을 잡아야 한다.

검사는 두가지 방식으로 분류해 볼 수 있다. maximum performance가 전제되는 검사 즉 최대의 노력을 해서 시험에 임한다 하는 것이 전제되는 검사와 자기의 있는 그대로를 반응한다는 habitual response를 전제로 하는 검사를 볼 수 있다. 지식과 기술평가는 maximum performance에 속하며 태도검사라는 것은 그런 식으로 할 수 없다. 즉 근본적으로 그 학생을 평가하는데 쓰는 것이면 태도검사는 될 수 없다. 태도검사는 있는 그대로를 알아 보는데 목적이 있는 것이다. 우리가 교육한 결과 학생들의 태도가 바뀌었더라 하는 program을 평가하기 위해서 쓰이면 몰라도 그 학생 자체를 평가하기 위해서 쓴다면 태도검사는 될 수 없다. 왜냐하면 사람은 윤리적인 판단능력이 있고 그것이 교육을 많이 받고 두뇌가 좋은 사람일수록 판단할 줄 알게 된다. 따라서 그런 식으로 태도검사를 하게 되면 학생 한명 한명의 평가를 위한 지식검사가 되고 만다. 「아! 이렇게 얘기하면 사회적으로 받아들여지고 윤리적으로 옳은 것이니까 그 방향으로 이야기하면 되지」 하고 자기 있는 그대로를 내놓을 수가 없게 된다. 따라서 지식검사가 된다.

와이드 유니어라는 사람이 The organization of man이라는 책에서 「개인은 현대사회에서 조직의 톱니바퀴처럼 전락하고 만다, 소외되고 만다」라고 하였다. 그 책에 How to cheat on Personality test 즉 인성검사에 있어서 어떻게 속임수를 쓰느냐? 예를 들어 회사원 취직시험에서 인성검사를 한다. 이때 회사측에서는 이 시험은 합격여부에 아무 관계없으니 있는 그대로의 여러분의 평소생각 행동방식 습관등을 적어 주십시오 하여 대상자들을 있는 그대로를 표현하게 하고 이 인성검사 결과 합격여부와 인사관리를 한다.

인사배치의 입장에서 인성검사를 쓸 필요는 많이 있다. 그러나 와이드 유니어는 말하기를 “이때 속지 말라”는 것이다. 다시 말하면 인성검사를 받을때 속임수를 써야지 솔직하게 쓰면 안될 수도 있는 것이다.

근본적으로 태도검사라는 것은 그러한 개인 평가는 될 수 없고, 있는 그대로를 반영시키고 그 결과를 가지고 우리의 교육 Program이 과연 간호원들에게 친절성, 사교성을 키울려고 했는데 그것이 키워졌는가를 보아야지 한 개인이 사교성이 있다 없음을 평가하는데 있어서 이 태도검사는 본래의 취지에 맞지않게 되어 있다.

평가의 원리에 대해 개론을 이야기하면 어떤 검사가 좋은 검사나 하는 검사의 양호도를 검토하는 기준은 타당도, 신뢰도, 객관도, 포괄도, 변별도, 실용도가 있다. 검사에서 가장 중요한 것이 타당도이다. 간호원 국가시험을 출제한다고 하면 과연 간호원의 전문적인 자질을 측정해 놓을 수 있는 옳은 시험문제가 되느냐는 것이 가장 중요하다. 신뢰도에 대하여 이야기하면 시험은 측정이니만큼 시험을 치렀을때 계속해서 똑같은 결과가 나와야 한다. 이것은 그 검사자체가 믿을 수 있느냐이지 학생 자신이 정직하느냐 아니냐의 뜻은 아니다. 이러한 이유로 객관식 출제를 자주 하게 된다. 그 다음 객관도에 대해서 이야기하면 객관도와 신뢰도라는 말은 사실 같은 개념인데 척도 자체가 신뢰도나 아니냐 하는 것은 신뢰도에 속하는 문제이고 채점자가 신뢰할 수 있느냐 아니냐 하는 것은 객관도의 개념이 되겠다.

예를 들어 어떤 A라는 사람이 채점을 하면 100점 반점에서 70점이 되고 B라는 사람이 채점을 하게 되면 30점이 되고 C라는 사람이 채점하면 50점이 된다면 이것은 문제가 되는 것이다.

그 다음 포괄도에 있어서는 이렇다 간호원으로서의 전문적인 지식을 갖고 있느냐 없느냐를 따지는 때는 상당히 광범위하게 퍼져야 하는데 어느 일부내용만 시험되어졌다면 어떤 학생이 있어서, 전체적으로 광범위하게 골고루 시험을 쳤을 때 점수가 좋은 사람이 어느 일부만 출제되어 실력을 했다면 여기에 객관식 검사의 장점의 불가피성이 여기에 있는 것이다.

예비고사에는 성적이 상당히 좋은 사람이 각 대학의 주관식 출제에서 실패된 학생들이 있다. 그것은 왜냐하면 주관식 출제는 출제자가 출제를 많이 낼 수 없으니까 자연히 몇개만 내게 된다. 자기는 이만큼 검토된 가운데 나는 이쪽이 나오지 않겠는가 하고 시험준비를 했는데 대학에서는 다른 쪽을 낸다면 헛다리를 짚게 되는

것이다. 그런 면에서 시험문제는 골고루 광범위한 내용을 잘 cover하지 않으면 이것이 좋은 시험이라고 얘기할 수 없다.

그 다음 변별도(discrimination)에 대해 얘기하면 시험이란 아는 사람과 모르는 사람과를 구별하는 것이 목적이다. 그것이 순위를 정하는 목적이든지, 일정수준을 이해하느냐 못하느냐 여하간에 우리가 우열을 가려낸다고 하는지 모르는지를 가려낸다고 하는지 시험의 목적이 있는 것이다. 시험을 쳐보거나 안쳐보거나 마찬가지라면 그것은 좋지 않은 것이다. 그러면 어떻게 변별도를 높히느냐 하는 것이다. 예를 들어서 순위고사(선발고사) 즉 norm referenced evaluation의 경우에는 100명이 시험을 쳤다면 100점에서부터 0점까지가 골고루 나와야 된다. 그것이 정상분포곡선으로 그려진다면 더 바람직할 것이다. 그러나 시험을 쳐보면 대체로 그렇지 않다. 보통 어느 쪽으로든지 기울어진다. positive square든지 negative square든지간에 기울어질 가능성이 참 많다. 기울어지든 안기울어지든 이 Coverage가 넓으면 넓을수록 변별하는 폭이 넓으면 넓을수록 좋은 시험이라고 얘기할 수 있다.

다음 실용성의 문제를 얘기하기로 하자. 우리가 왜 객관식 출제를 하게 되느냐 논문식으로 한다면 그 채점시간이 많이 걸린다. 시간이 첫째 문제이다. 따라서 채점자에 대한 수당이 많이 들게 되고 비용, 노력도 많이 든다. 그러나 객관식 출제를 하게 된다면 예를 들어 40만명의 수험생에게 5과목의 시험을 치룬다면 그것을 과학적으로 computer 처리를 하면 10일이면 끝낼 수 있는데 240만에 가까운 것을 주관식으로 채점을 해야 한다면 1년내내 해도 어려울 것이다. 그래서 주관식으로 못하는 이유가 거기에 있는 것이다. mass testing에서는 불가피 객관식으로 하게 된다. 간호원 3,000명에 대해서도 마찬가지로 주관식으로 가미하기는 도저히 불가능하다. 그것은 출제자도 많이 동원될 것이고 채점인원 수당도 문제가 된다. 그러니까 시험은 아무리 이상적으로 짜여져 있다 해도 현실이니까 현실적으로 시간, 경비, 노력이 적게 드는 것일수록 좋은 시험이라고 하는 외형적인 조건이지 본질적인 조건은 안되지만 그런 조건을 무시할 수 없다. 맨처음 우리가 생각할 수 있는 것은 일종의 피상적인 타당도이지만 그것은 “간호원에

대한 시험같다”라고 되어야 한다(face validity) 결국은 모토가도 서울만 가면 된다. 간호원에 대한 시험같다. 그렇지 않다, 무슨 관계가 있느냐 하지만 이 시험이 공신력을 갖는다면 이해가 부족한 사람들에게는 역시 간호전문직에 속하는 또는 전문직의 판단에도 그것이다. 그러니까 여러분들이 시험문제 내는 것은 간호학자로서 간호전문직에 관계되는 시험문제인 것 같다고 하는 인상을 줄 수 있는 시험문제를 낸다는 것이 첫째로 중요한 얘기다. 이것은 본질적인 것은 되지 못한다.

그 다음 중요한 것으로는 내용파악이 되겠는데(content validity)이다. 이것은 아주 본질적이다. 이러한 학과시험 간호학이면 간호학과 시험에서 가장 중요한 것이 내용타당도라고 얘기할 수 있다. 그러니까 먼저 말한대로 간호학의 각 분야별로 학과 전문가들이 출제를 해야 하고 내용의 타당도 속에서는 포괄도도같이 문제가 된다고 보면 된다.

골고루 내용을 Cover했느냐 안했느냐를 같이 포함한다고 볼 수 있다. 그 다음에는 실험적인 타당도로서 경험타당도라고도 이야기 하지만 실제로 그 시험에서 성적이 좋은 사람이 간호원으로서 전문직 수행에 있어서 좋은 성과를 내더라도, 시험을 치워서 무엇을 하느냐 결국은 간호원의 국가시험을 친다고 하면 간호원으로서 전문직 수행을 잘할 것이라고 내다보는 또 그렇게 증거를 부릴 수 있는 시험이 될 것이다. 이 경험적 타당도 실험적 타당도 속에는 예언타당도도 포함된다. 그 시험에서 좋은 사람들이 5~10년 지나가면서 훌륭한 간호원으로서 업무수행을 하드라고 포함되기도 한다. 그 다음에 우선 타당도에 대한 얘기도 해야 한다. 과연 그것이 여러분들이 국가시험을 치루는다는 직접적으로는 관계되지 않기 때문에 우선 타당도 같은 것은 심리학적인 다른 검사를 할적에 많이 쓴다. 지능검사에도 문제가 된다.

factorial analysis의 factor를 세어서 과연 지능의 구성요인을 제대로 cover했느냐 못했느냐를 따지는 문제가 된다. 내용타당도도 제일 중요한 문제가 되는데 이 내용타당도라는 것은 학과 전문가가 따지게 된다. 이 내용타당도를 따지는 기준이라는 것은 쉽게 말하면 교수요목(Syllabus)이라고 볼 수 있다. 이것은 여러분들이 말하자면 미리 합의를 보아야 되겠다. 각

교수의 제량에 속하는 것보다는 간호학회, 또는 간호협회가 전문가 몇사람을 위촉해서 내과간호학, 소아과간호학은 이런이런 것을 알아야 한다. 그런 식으로 말하자면 적어도 이런 내용은 cover해야 한다는 것을 각개인 교수로서 그 학문의 권위자가 만들어야 하지만 교수요목이라는 것이 그렇게 쉽게 객관적으로 설정하기는 어려운 것이니까 이것은 협의를 거쳐 가지고 여러분들이 교수요목을 만들어야 할 것이다. 아마 오늘 내일에 있는 역삼에 있어서 이 문제가 일차적으로 여러분들이 해야 할 일에 속한다고 생각한다. 종전에 되어 있다면 할 필요가 없겠지만 있다 하더라도 재검토를 해야 한다. 과연 과목이 몇개인지 과목별로 먼저를 작업이 무엇을 가르쳐야 할 것이냐 무엇을 가르쳐야만 훌륭한 간호원이 되느냐 따라서 무엇을 출제할 것이냐를 일단 검토하는 것이 되어야 한다. 이것이 제대로 안되면 전문직으로서의 간호원의 면허시험을 치는 내용이 편파적으로 되거나 피상적으로 되거나 또는 시대착오적인 옛날 구식으로 켜다고 할 수 있다. 새로운 간호학의 발전추세를 비쳐보아 적어도 소아과 간호학은 이런것 내과간호학은 이런이런 것을 알아야 되겠다. 예컨대 이런 것을 분과별로 다루어서 얘기해야 한다. 그리고 내용타당도가 중요한데 이것은 출제자가 학과에 전문가라고 하는 입장에서 주관적으로 판단할 수 밖에 없는 것이라고 그것이 이번과 같은 협의회에서 이렇게 이렇게 이루어 내용구성은 이렇게 하자 하고 거기서 출제는 이렇게 하자 하는 협의가 진행되므로써 타당도는 여러분들이 해야 할 일이 되지 않겠는가.

여러분들이 해야 할 작업이 구체적으로는 이 교육목표의 이원분류표를 작성하면 될 것이다. 다섯분과면 다섯분과 나름대로 우리가 내용의 측면에서 어떻게 할 것이냐 그것을 어떻게 세분화 해 갈 것이냐 둘째 문제는 어떤 문제를 낼 것이냐 그것을 만들어 놓고 동시에 그것을 어느 행동으로 이해를 시킬려고 하느냐 기본개념의 definition을 기억할 수 있느냐 기억하느냐, 못하느냐 이것은 Broom의 이론이라고 하면 지식이 되는 것이다. 그것은 기초개념의 정의를 내릴 수 있는 기본개념의 definition을 갈 수 있느냐 없느냐 하는 상당히 중요한 것이 된다. 기본개념의 이해는 기초적인 것이다. 그 다음은 그것을 자기나름대로 이해했느냐 그러니까 기계

적으로 기억하는 것이 아니라 그것을 자기방식으로 다시 조서할 수 있느냐 이런 것이 공통된 것이다. 그다음 그 기본개념을 배운 것을 가지고 어떤 새로운 사태에 적용해 볼 수 있는 힘이 있느냐 이렇게 보면 여기서부터 Higher mental Process한다고 한다. 이것을 어떠한 새로운 사태에 적용을 한다고 할 것 같으면 얼마큼 분석을 제대로 해 놓았느냐 여러가지 사태적인 그 Situation factor가 여러가지로 있겠는데 그것을 어떻게 종합할 수 있느냐 다시 평가할 수 있느냐 그것을 위해서 여러분들에게 Broom의 이론을 참고해서 봐야 할 필요가 있지 않겠느냐고 본다. Broom의 가장 큰 공헌이라고 볼 수 있다. 이런 이용표를 놓고 여러분들이 이것은 이해선으로 놓아야 하겠다. 적용선에 놓아야 하겠다. 하는 A라는 개념에서는 더 이런 측면에 놓아야 하겠다는 weigh를 정해야 한다. 이 개념에 대해서는 상당히 고차적인 이해 분석능력까지도 주어야 하겠다. 기대해야 하겠다. 측정해야 되겠다는 것을 여러분들이 놓아야 한다.

문제를 출제할 때 그만큼 전문적인 자격고사의 출제는 하루에 5~10개 밖에는 만들 수 없다. 단 출제는 이런 기본적으로 2원분류를 해서 작업방향을 제시해 놓고 여기에 관계되는 것은 다음과 같은 2원분류표에 의하여 뽑아놓고 만들어야 한다.

2원 분류표에 따라서 각 item idea를 3배수를 뽑아 놓아야 한다. 그것을 다시 이원 분류표와 비교해 보고, 문제가 얼마큼 참신한가를 내가 출제할 가능성이 있느냐 없느냐에 의해서 문제수를 뽑아야 한다. 참고자료로 이야기 하자면 이런 분류표를 만들어 놓았다고 해서 꼭 따추라는 이야기가 아니라 이런 분류표에 맞도록 노력하라는 이야기다. 결국 고급전문직이란 그것을 Higher mental process를 작용해야 한다. 고등한 사고능력이 정해지지 않으면 심화된 이해를 하지 못하면 그것은 전문적이 아니다. 그런 점에서는 Higher mental Process를 제지 않으면 즉 적용능력 분석능력 종합능력을 제지 않는 시험이라면 그는 전문직 자격고사라 할 수 없다. 지능검사, 학력검사, 적성검사 등 시험문제를 출제할 때 많이 만드는 것이 좋은 것이 아니고 독창적인 좋은 문제를 만들려면 하루에 multiple choice 5개 밖에 만들지 못한다. 다른 문제를 베끼는 식으로 출제를 해서 많이 만드

는 것이 좋은 것이 아니고, 독창적인 좋은 문제를 만들려면 많은 시간이 필요한 것이다. 대체로 item idea를 3배수로 잡고 그리고 단독출제는 언제든지 나쁘다 왜냐하면 그것은 혼자 작업한다는 것은 Creative-Productive는 하지 않는 것이 있기 때문이다. 또 하나는 오답지를 잘 만들어야 하는데 이것은 idea가 많아야 한다. 그래서 출제는 자자 책임을 지고 하더라도 같이 작업을 하면 더 질이 좋고 능률적으로 된다. 쉽게 말하면 설문항 질문항은 먼저 써서 써야 한다. 일반적인 원칙은 질문항은 분명하게 이해될 수 있도록 하며 질문항은 짧고 답지항은 긴 것이 좋다. 질문항은 길고 답지항이 간단하면 길이 있는 이해가 되지 않는다. 흔히 상상적인 체제에서 질문항이 길고 답지항이 짧은 경우가 많은데 이것은 길이있는 이해가 안되고 가려내는 데 목적이 있어서 좋지 않다. 결국 item system은 반드시 짧아야 되는 것은 아니지만 답지를 만들 때는 답지의 길이는 비슷해야 한다. 일반적으로 어떤 패턴이 생기나 하면 정답이 될려면 필요조건 충분조건을 다 갖추어야 하므로 길어지고 오답지는 짧아진다. 그러면 이 추측을 할 적에 모든 경우 긴것을 고르면 대체로 맞는다는 요령이 생긴다. 그런데 그러한 추측의 가능성을 약화, 예방하기 위해서는 답지의 길이를 다같이 해야 한다. 그럼 어떻게 오답지를 작성해야 하느냐, 답지를 만들 때에는 먼저 오답지부터 만드는 것이 좋다. 결국 이 객관식문제에서 승패를 좌우하는 것은 오답지를 잘 만든느냐, 못 만든느냐에 달려있다. 가능하면 오답지를 먼저 만들어 놓고 그리고 정답지를 작성하는 요령이 대체로 출제성과를 올릴 수 있는 방법이 될 수 있다고 말할 수 있다. 그밖에 객관식 출제에는 경험으로 봐서 오답지를 만드는데 생명이 있다. 그 다음에 출제방식에 있어서 또 한가지는 일문일답되는 것은 좋은 것이 될 가능성이 약하다. 그것은 역시 기억을 해서 기억적인 것이나 가벼운 이해는 되더라도 적용은 되지 못하고 고차원적인 사고능력은 되지 못한다. 그러니까 간호사 때에 있어서 Problem Situation을 기술해 놓고 그것을 출제한다는 idea를 넣어서 종합표를 만들어야 한다. 또는 가장 쉬운 것은 통계표를 만들어 놓고 가설적인 상황을 놓고 문제를 거기에서부터 문제를 만들면 된다. 여기에 이것은 어떤 개념이 적용된 처리법이나, 이러한 식으로

물론 방식으로 하면 된다. 한 Problem set을 내놓고 거기서 대체로 5~7쪽 문제를 빼놓을 수 있도록 단들던 가장 훌륭한 출제가 된다. 이러한 Problem situation test를 만들 수 있는 능력은 상당한 시간이 걸린다. 왜냐하면 Problem 사태를 고안해 내야 하기 때문이다. 결국 간호업무라는 것은 일상 교차적인 두뇌작용도 그렇지만 그것을 해결해야 할 능력을 요구해야 되지 않겠는가 라고 나는 생각하기 때문에 간호상황에서의 문제되는 상황을 출제해야 한다. 이것은 강의실에서 배운 것을 새로운 사태에 적용해서 이해해 보는 능력향상의 방향으로 제시될 수 있다.

답하는 요령이 새로우면 생각 자체도 새로울 수밖에 없다.

말을 바꾸면 출제형식을 바꾼다는 것 자체는 사고형식을 다르게 요구한다는 것이다. 예컨대 출제방식에서 또 하나의 방식은 Set response를 만들어 놓고 답지를 이런 식으로 만드는 것이다. 다음의 글에는 갑, 을 두가지의 statement들이 있다. 그것을 읽고 만약 갑이 을의 원인이 되던 1번에 을이 갑의 원인이 되던 2번에 갑과 을 두가지다 타당한 statement이면 관계가 없으면 3번에, 갑이나 을 가운데 어느 하나가 force statement다 일 때는 4번에 위와같은 상황을 놓고 하면 문제만들기는 상당히 쉽다. 이때 학생들은 생각 안할 수밖에 없는 것이다. 갑을 statement가 틀리나 옳으나 어느것 하나가 틀린 것이나를 생각하게 된다. 이 방식은 관계분석의 출제방식이다. 관계를 규명하는 것이니까 그만큼 Higher mental Process가 작용할 수 있다. 이것은 참으로 어렵지만 단들면 참 재미있는 것이다. 답하는 학생은 하나의 상황이 일어났으면 거기에 복합된 요인이 작용하는데 그 인과관계가 무엇인가를 규명해야 문제를 풀 수 있게 되는 것이다.

요컨대 문제 출제방식에서 “결핵유병율은 얼마인가 쓰라” 이런 방식은 가장 초보적인 기본적인 개념으로 Higher mental Process가 작용하지 않을 가능성이 많으므로 문제상황을 놓고 그것을 뽑아내면 그 상황에서 종합적인 판단 속에서 답을 해야 되기 때문에 그만큼 종합적인 사고를 요구하게 된다.

시험 후에는 문항분석을 해야 한다. 전년도 간호원국시 응시자의 몇백 학생의 답지를 뽑아

서 문항분석을 한다.

문항분석(item analysis)은 어떻게 하는 것이 나?

첫째, 문제가 얼마큼 어려웠느냐? 대체로 정답율로 보아서 80%를 넘으면 그 문제는 쉬어서 사용할 수 없다. 20~30%선이면 너무 어렵거나 추측요인에 의해 작용할 가능성이 많기 때문에 사용하지 않는다.

간호원 국가시험은 전문적인 자격시험이니까 시험문제는 분석해 보아서 50% 이상이 되면 그것은 좋은 문제가 되지 못했다. 너무 쉬었다는 것을 나타낸다.

둘째, 답지의 반응분포를 본다.

예를 들어 답지가 4개 있는데 1번 답지가 80% 2번 답지 20% 들어 갔다면 이것은 4선택지의 문제가 아니라 양자택일이 되는 것이다. 진위형의 문제가 되는 것이다. 다시 말하면 3번, 4번 답지를 유혹자로서 모든 사람의 혼동에 빠뜨리는 기능을 전혀 못한 답지이다. 제대로 풀려면 정상분포가 되어야 한다. 이러한 문항분석을 어느 기관에서 하여야 될지 모르지만 보건원의 협조가 가능하다면 일년에 한번씩은 주기적으로 하여 item pool을 만들때에 이러한 답지

의 반응분포를 보아 골고루 펼쳐있는 것을 item pool에 집어 넣도록 하는 방식으로 하고 그것을 feed back할 필요가 있다. 그 다음 과제는 문항에 얼마큼 discriminating power가 있느냐이다(discriminating power index).

변별도 이것은 쉽게 말하면 좋은 문제라고 할 것 같으면 머리가 좋은 상위집단에서 아이들은 다 pass하고 하위집단은 다 실패한다고 할때 상하로 갈라내는 변별력을 가려낼 수 있는 힘을 가졌다고 할 수 있다. 이때 상하집단을 무엇으로 나누느냐, 그것을 자체검사로 나눌때는 예를 들어 100문제를 쳐서 점수좋은 아이들은 상, 점수가 나쁜 아이들은 하로 나누어서 절반으로 나누면 그것은 item deserminating index라고 얘기한다. 이 변별이 바람직하느냐 아니냐는 알 수 없다. 왜냐하면 총점이 얼마인지 모르기 때문이다.

Item discriminating index나 item false index하는 것은 위의 기준으로 따진다. 이렇게 문항을 실증적으로 검토해 봐야 출제기술이 늘게 되고 또 출제가 보장된 질의 것이 될 수 있는 길이 된다.