

공직 적격성 검사의 개발

김 청 택* 김 명 소** 이 순 목***

요 약

이 연구의 목적은 국가 고등고시 1차 시험에서 우수한 인재를 선발하기 위하여 적성검사를 개발하는 것이었다. 일반적성이론에 관한 문헌조사와 재직자와의 면접을 통한 직무분석 결과에 기초하여 언어능력, 상황판단능력, 자료해석 능력, 상식의 네 가지 하위영역을 측정하는 검사를 제작하여 대학생 집단, 재직자 집단, 수습사무관 집단을 대상으로 1차 실험평가를 실시하였다. 문항분석을 통해 문항의 양호도를 검토한 후 이러한 정보를 기초로 새로운 문항을 제작하고 대학생 집단, 재직자 집단, 수습사무관 집단을 대상으로 2차 실험평가를 실시하였다. 최종적으로 개발된 문항들은 신뢰도와 타당도 면에서 수용할만한 수준이었다. 개발된 공직 적격성 검사의 의의와 앞으로 고등고시에 지원할 응시자들에게 미치는 순기능을 논의하였다.

Abstract

The purpose of this research was to develop Public Service Aptitude Test(PSAT) in order to select high performer in the public sector of the government. Based on both literature review and job analysis using the interview with job incumbents, the test which measured verbal ability, situational judgment ability, data interpretation ability, and knowledge on common sense was prepared. The items was administered in undergraduate students, job incumbents, and trainees of public service officers to obtain information about the goodness and appropriateness of items. Base on the results of the item analysis, new items were made and administered in another undergraduate students, job incumbents, and trainees of public service officers to select the final test items. The reliability and validity of the developed PSAT were acceptable level. Finally, the impact of the PSAT on potential applicants and future suggestions were discussed.

*서울대학교 심리학과

**호서대학교 산업심리학과

***성균관대학교 심리학과)

최근에 조직이 처한 외부환경이 급속하게 변화하고 새로운 테크놀로지가 개발됨에 따라 일의 세계에도 많은 변화가 이루어지고 있다. 일을 수행하는 방식에도 많은 변화가 일어나고 있고 조직이 구성원들에게 요구하는 능력이나 기술 등도 과거와는 많이 달라지고 있다. 예를 들어, 직무와 직무간에 분명한 경계가 없어지고, 팀으로 수행하는 일이 많아지고 있으며, 변화하는 일의 세계에 효과적으로 적응하기 위하여 조직 구성원 개개인의 학습능력이나 적응력이 강조되고 있다. 이러한 추세에 따라 조직이 새로운 구성원을 채용할 때 중요하게 여기는 요소나 채용할 때 사용하는 도구와 방식에서도 많은 변화가 이루어지고 있다.

Schmitt와 Chan(1998)은 테크놀로지의 급속한 발달, 작업수행에서 팀제의 증가, 의사소통 기술의 발달, 세계화, 고객만족에 대한 관심의 부각과 같은 다섯 가지 중요한 변화요인이 조직장면에서의 인사선발에 영향을 미칠 것이라고 주장하였다. 최근에 직무수행에 요구되는 지식이나 기술의 수명이 점차 짧아지고 지속적으로 새로운 지식이나 기술이 출현하기 때문에 앞으로 선발장면에서는 특정분야에서 요구되는 이미 습득된 '세부적인(specific) 지식이나 기술'보다는 새로운 지식이나 기술을 빨리 습득할 수 있는 '일반적인 학습능력(general learning ability)'이 더 중요해질 것이다. 따라서 선발장면에서 특정 지식이나 기술을 측정하는 검사보다는 일반적인 학습능력을 측정할 수 있는 검사의 사용이 증가할 것이다. 국내에서도 기업체의 입사시험에서 전공시험을 보지 않고 적성검사를 실시하는 기업이 증가하는 것도 이러한 추세를 반영하고 있다고 볼 수 있다.

이 연구에서는 국가 고등고시 1차 시험에서 기본적 능력을 측정하는 적성검사를 개발하는 절차와 결과를 제시하고 있다. 공직 적격성이란 공직 및 그에 관련된 업무를 잘 수행할 수 있는 능력인데 이에 대한 측정용 정의가 현재는 지식중심으로 되어있다. 즉, 1차 2차 시험이 모두 지식을 묻는 시험이고 3차라고 할 수 있는 면접은 실질적 역할을 못하고 있다. 기존의 1차 시험과목들은 "암기와 단편적 지식측정위주"(중앙인사위원회, 2000, p.1)로 검사함으로써, 변화하는 시대적 환경에서 국가경영에 참여하는 유동적이고 적응적인 인재를 발굴하기에는 부족하였다. 또한 기존의 시험들은 순발력있는 암기와 단기간의 집중으로 고득점을 올릴 수 있다는데서 대학사회를 떠나없이 국가고시 준비장소로 전락시키는 파행교육의 부분적 원인으로 비판 받아왔다. 이러한 고시제도는 정보화사회 혹은 지식기반사회 등으로 특정 지워지는 현대사회가 요구하는 공직 적격성을 측정하기에는 한계를 지닌다. 현대의 과학기술의 급격한 발전과 정치·사회·문화의 패러다임의 급속한 변화는 공직지원자들이 공직자 세계의 지식기반을完비한 상태에서 채용되는 것을 어렵게도 하지만 바람직하지도 않게 만들고 있다.

개발된 공직 적격성 검사(PSAT: Public Service Aptitude Test)에서 측정되는 문제해결능력(언어능력, 자료해석력), 상황판단력 및 기본소양(사회상식)은 기존의 국가고시 1차 시험에 포함되었던 헌법, 한국사 및 기타 전문지식과목에서는 측정될 수 없는 내용들이다. 문제해결능력 과 상황판단력이 기본적 잠재력이라면 기본소양은 기본적 성취도라고 할 수 있다. "기본적"

성취도를 보는 경우에 암기의 요소가 완벽하게 제거되는 것은 아니지만, 적어도 과거와 같은 과도한 암기위주의 대학교육과 암기위주의 고시생 양성이라는 파행에서 벗어나는데 일조를 하게 된다. 또한 잠재력을 본다는 것은 공직사회의 초급관리자로서 적응하고 성장하기 위한 기본인지능력과 상황판단력을 중심으로 하므로, 특별히 수험생들이 암기를 통해 준비해야할 필요성을 감소시키는 특징을 지닌다. 기본인지능력은 수험준비로서 책 몇 권 더 읽었다고 단기간에 증가하는 것이 아니며 상황판단능력 역시 암기위주의 집중적인 수험준비가 전혀 통할 수 없는 측정방식이다. 따라서 과도한 고시준비로부터 탈피하는데 일조를 하면서, 기본적 성이 있는 지원자들이 과거와 같이 많은 준비 없이도 쉽게 응시하도록 고무함으로써 우수한 능력을 더욱더 많이 공직에 유인하는 목적에도 기여하게 된다.

하위 영역 설정

1. 하위 영역의 설정의 기준

공직 적격성은 다양한 인지적 능력으로 구성되어 있으므로 이 공직 적격성의 모든 측면을 측정하는 것은 가능하지 않다. 본 검사에서는 다음과 같은 원칙 하에 공직 적격성의 영역을 한정하였다.

첫째, 고시에서 2차 시험이 전문적인 지식의 정도를 측정하고 있으므로 공직 적격성 검사에서는 초급관리자로서 필요한 기본적 소양을 측정하는 일반 적성과 공직관련 적성을 측정한다.

둘째, 현재 직무에 대한 지식이나 기술보다는 인지적인 문제해결능력, 환경 적응적인 상황판단력과 같은 초급 관리자로서 필요한 잠재력을 측정한다. 이러한 능력들은 기존의 국가고시 1차 시험에 포함되었던 헌법, 한국사 및 기타 전문지식 과목에서는 측정될 수 없는 내용들이다.

셋째, 이미 개발되어 사용되고 있는 검사에서 타당성이 검증된 영역을 참조점으로 하여 영역을 구성한다.

공직 적격성 검사는 언어능력, 상황판단능력, 자료해석 능력, 상식의 4가지 영역의 능력이 측정되었는데 이는 일반적성이론과 직무분석의 결과에 근거한 것이었다. 이 영역 중 언어능력과 자료해석 능력, 상황판단력은 Gardner(1983)의 지적능력에 대한 다면이론에서 제안된 7개의 영역 중에서 공직자와 유관하다고 판단되는 언어능력과 논리 수리능력의 영역에서 도출되었다. Gardner의 지적능력 중 타인의 기분, 기질, 동기, 의도를 파악하고 구분 짓는 능력인 개인간 지적능력은 현 한국사회의 현실상 인성으로 오해를 유발할 소지가 있어서 본 검사에 포함되지 않았다.

또한 Sternberg(1995)의 실무적 지능에 근거하여 실무와 관련된 영역을 추출하였다. Sternberg에 의하면, 실무적 지능은 다음과 같이 크게 세 가지로 나누어 볼 수 있다. 첫째 대인관계적 혹은 사회적 지능은 자신을 포함한 사람들의 정서, 사고, 행동들을 이해하고 이를 바탕으로 자신을 관리하거나 타인의 언행을 예측하고 적절하게 대응하는 능력이다. 둘째, 상황적 혹은 업무적 지능은 집단 장면에서 주어진 상황을 정확히 판단하고 해결책을 제시하는 능력, 주어진 상황을 개선하기 위해 방해가 되는 긴장이나 갈등의 요소를 해결하는 능력이다. 마지막으로 상식적 지능이란, 업무수행에 필요한 기본적인 지식과 정보, 사회의 규범이나 약속을 이해하고 따르는 행동 및 태도(예절/매너) 등을 말한다. 위의 세 가지 실무적 지능 중, 대인관계적 혹은 사회적 지능은 인성과 관련된 요인이어서 지필검사로 측정하는 데는 많은 제한이 있다. 이러한 제한을 감수하면서 PSAT에 포함시킨다면 법적으로 혹은 도덕적으로 여러 가지 문제를 야기할 가능성이 있기 때문에 제외하였고, 한편 상식의 영역은 지필검사가 가능하므로 PSAT에 포함되었다. 상황적 혹은 업무적 지능은 직무수행과 직접 관련된 적성으로서 중요한 역할을 하고 있다. 연구의 초기에는 상황판단능력에서 소위 실용적 지능(practical intelligence)을 측정할 계획이었다. 실용적 지능의 측정은 업무상에서 일어나는 특정 상황을 피검자에게 주고 그 해결책을 찾아내게 하는 방식을 취하게 된다. 이때 경험적 정답(empirical key) 방식, 즉, 정답을 실제 업무에 종사하는 사람이 합리적으로 생각하는 해결책을 정답으로 삼는 방식을 택하게 되는데, 정답에 대한 이견이 발생할 소지가 있어서 현행 고시 제도에서는 반영하기 힘들다는 결론에 따라, 실용지능보다는 논리력, 추리력, 문제해결 능력으로 정의되었다. 따라서 이 영역은 일반 적성과 실무적성이 중복되어 있다. 이렇게 적성이론에 근거하고 현실적 문제를 고려하여 최종적으로 선택된 네 가지 영역은 경험적 연구인 검사 개발용 직무 분석에 의하여 각 영역별 타당성이 검증되었다.

2. 필요역량에 대한 조사

(1) 면접 방법

행정고시 및 기술고시를 통해 임용된 자로서 임용 후 2-3년이 경과한 사람가운데 21개 직렬 18개 부처에서 52명을 선정하여 면접하였다. 5명의 연구원들(조직심리학 박사 3명, 심리측정 박사 2명)이 각각 피면접자 2인을 1조로 하여 약 1시간 동안 반구조화된 면접을 진행하였다. 면접의 목적은 공직 초급관리자(5급 공무원)에게 요구되는 능력(인지/비인지)과 지식을 파악하는 것이었으며 이때 특정 부서에서 요구되는 능력과 지식을 묻기보다는 5급 공무원의 직무수행에 전반적으로(공통적으로) 필요한 능력과 지식을 조사하였다.

(2) 면접 결과

공직자로서의 업무를 수행하는 데 필요한 능력으로 피면접자가 응답한 내용은 크게 언어능력, 논리력 및 판단력, 자료처리 및 수리능력, 그리고 일반지식 혹은 상식으로 나눌 수 있었

다. 구체적으로 언어능력에는 언어 이해력과 표현력, 의사소통 능력, 자신이 알고 있는 지식을 종합할 수 있는 능력이 포함되었고 자료처리 및 수리능력에는 자료분석 능력, 정보분석 능력, 많은 자료로부터 핵심적인 자료를 추출하는 능력이 포함되었다. 논리력 및 판단력에는 논리력, 문제해결 능력, 의사결정 능력, 기획력, 기존의 지식을 새로운 영역에 응용하는 능력이 포함되었다. 마지막으로 일반지식 및 상식에는 컴퓨터 관련 지식, 인터넷 지식, 법률지식, 법 적용 마인드, 국제감각 등이 포함되었다. 이 외에 언급된 내용은 창의력, 새로운 것에 대한 학습능력, 종합적 사고력, 분석력 등이 있었다.

검사개발의 논리

1. 검사개발의 원칙

본 검사를 개발함에 있어서 검사이론에서 요구되는 기술적 수준을 유지하는 것과 더불어서 공직에서의 초급관리자를 위한 검사라는 점을 고려하여 다음과 같은 다섯 가지 원칙에 의해 검사가 개발되었다.

첫째, 공직 초급관리자의 학습잠재력과 기본소양을 중심으로 공직 적격성을 평가하는 검사로 전문지식의 성취도검사를 지양하였다. 전문지식에 대한 성취도 검사는 2차 시험에서 평가되는 점을 감안하여 1차 시험에서는 공직 초급관리자로서 필요한 잠재력과 기본소양을 다면적이고 체계적으로 평가할 수 있도록 검사를 개발하였다. 이미 국내 대기업(삼성 SSAT, LG FAST, 선경 종합적성검사 등)에서 직무적성검사라는 제목으로 실시되어온 검사들과 외국(예: 일본의 1종 국가공직지 행정구분 교양시험, 영국의 속진 임용제도의 1차 시험)에서 실시되고 있는 공직자 임용시험을 볼 때, 우리 나라에서도 이러한 평가방식을 도입하는 것은 시대적 추세에 부응하는 조치로 볼 수 있다.

둘째, 단순평가가 아닌 복합적 문제방식을 지향하였다. 종래의 1차 시험이 암기력을 평가하는 단순평가 위주라고 한다면 이번에 개발될 문제는 잠재력과 기본소양을 학제적·복합적으로 평가할 수 있는 문제방식이라고 할 수 있다. 즉 문제의 지문도 보다 길어지고, 많은 인지적 사고력을 요하는 상황을 제시하여 단순암기력과 무관한 기본능력과 소양을 검사하는 방식으로 출제되었다.

셋째, 보다 현실에 가까운 문제 구성을 추구하였다. 공직에서 다양한 종류의 문서에 대한 해독 및 추론이 요구되는 것을 감안할 때, 언어 능력의 측정에 사용되는 지문들은 보다 길고 많은 인지적 능력이 요구되는, 다양한 내용들을 포함하였다. 자료해석영역에서는 행정 현실에서 다루는 실제 문제(예: 물가변동, 환율, 실업률, 인구증가, ...) 상황에서의 자료 및 통계해석 능력과 수치처리 및 추론능력이 검사되었다.

넷째, 지나치게 전문적이고 높은 지식의 문제는 지양하였다. 지식의 정도가 문제를 푸는데 도움을 줄 수는 있으나 문제풀이에 결정적인 요소가 되면 PSAT가 1차 시험으로서 원래 의도하고 있는 공직 적격성 측정 목적에 어긋나는 것이다.

마지막으로, 구체적이지는 않지만, 기본적인 대원칙으로 고시학원에서 학습될 수 있는 내용보다는 정상적인 교육을 통해 다양하게 학습할 수 있는 능력을 측정하려고 시도하였다.

2. 문항제작의 논리

4개 영역에서 문항들은 지원자 전체 중 10%-15%를 차지하는 합격자들이 평균 80%이상 이 되도록 한다는 준거를 가지고 제작되었다. 이는 검사사용자의 요청에 의해 정해진 준거이었다. 학습잠재력을 측정하는 언어영역, 상황판단영역 및 자료해석영역에서는 다양한 내용과 유형의 문항들이 제작되었지만, 특히 기본 성취도를 측정하는 상식영역에서는 고급공무원이 알아야 할 다양한 기본지식이라는 강한 외적 준거를 중심으로 한 검사가 제작되었다.

그러나, 각 영역별로 다양화 가능성의 폭은 동일하지 않다. 언어, 상황판단, 자료해석의 세 영역 중 언어영역은 언어이해 및 종합력이라는 넓은 범주에 일치하는 다양한 내용 및 유형을 바탕으로 문항제작이 용이하다. 반면 상황판단영역의 경우, 내용을 보면 주로 문제해결, 의사결정, 기획, 응용 등에서의 논리력을 중심으로 문항이 제작되는데 그런 문제 자체가 가지는 전형적인 유형이 있으므로 언어영역에서처럼 다양한 유형을 보이기는 어렵다.

자료해석영역은 내용으로 볼 때 자료분석능력과 필요정보 추출능력이 되며, 이들 역시 문제자체의 전형적인 유형이 있어서 언어영역에서처럼 다양한 유형이 되기는 어렵다. 상식영역에서는 가장 다양한 분야에서의 지식을 기초로 문항이 제작되므로 비교적 내용영역이 다양하게 된다. 유형으로 볼 때 정교한 기억과 상식추론의 유형이 되면서 상대적으로 복잡성이 덜 강조되는 측면이 있다. 즉 일면 단순하게 보일 수도 있다.

그러나 언어, 상황판단 및 자료해석의 경우는 복합적 문제방식이 강조되므로 문제의 지문이 길어지고 단순암기력보다는 기본능력과 소양을 요구하는 문제가 강조된다. 따라서, 난이도로 볼 때는 이 3개 영역이 상식에 비해서 높아진다. 상식영역은 성격상 '알고 모르고가 분명한' 영역으로서, 너무 세부적인 것을 문항화 하기보다는 "고급 공무원이 알아야 할" 지식이라는 준거를 중심으로 하여 제작되므로 비교적 난이도는 나머지 3개 영역에 비해서 높지 않을 수 있다.

3. 문항의 난이도 설정

검사개발을 위한 이론으로서 고전검사이론이 아닌 문항반응이론 (Baker, 1985 등)을 사용하였으므로 문항분석에서도 같은 논리를 적용하였다. 문항반응이론에서는 난이도라는 용어보다는 곤란도라는 용어를 선호하므로 그 관례를 따르기로 한다. 문항반응이론에서 문항의 곤란도는 응답자 모수의 측정체계 속에서 정의된다. 따라서 응시자의 능력수준은 어떤 수준의

문항으로 즉시 연결이 가능하고 ‘일정 수준의 사람들이 몇 점을 얻기 위해서 어떤 수준의 문항이 필요할지’에 대한 의사결정이 가능하다.

그렇다면 이 연구의 관점에서 응시자와 문항을 연결시켜보기로 한다. 이 연구에서 응시자들이 치르게 되는 PSAT 관점에서 능력수준을 분명하게 언급할 수 있는 사람들은 아직은 아무도 없다. 아직 검사가 실제 5급 공무원시험의 현장에서 실시된 적이 없기 때문이다. 따라서 1차 실험평가 자료와 2차 실험평가 자료간에 척도연결을 위해 수집한 E여대와 SK대학교 학생집단을 전체 응시자들의 능력척도에서 중간에 있는 집단으로 가정하는 방식을 사용하였다. 앞으로 이 집단을 가교(bridge)집단으로 부르겠다. 물론 이 가교집단의 능력이 미래의 응시자 모집단에서 어느 위치에 해당할지 아직은 명확하지 않다. 따라서 현재 개발되는 검사들 가교집단의 능력수준에 연결하는 것은, 검사개발의 목적상 꼭 필요하긴 하되 개략적 작업이라는 한계가 있다는 점이 언급되어야 할 것이다.

가교집단은 5급 국가고시에 대한 전체응시자의 능력척도에서 중앙에 위치할 것으로 추정되는 사람들이다. 즉 이 집단의 능력척도(능력분포는 문항반응이론에서의 가정에 따라 정규분포를 가정)에 있어서 위치(location) 모수와 표준편차의 모수를 표준정규분포에서 “일정한”값으로 부여하고, 이들에게 현재 개발중인 검사를 실시했을 때 문항의 곤란도 모수가 부여될 수 있다.

즉 가교집단의 능력의 위치(location) 모수 값을 0으로 하고 표준편차를 1로 주면 그에 대응하여 문항들의 곤란도가 능력 모수의 측정체계 속에서 산출된다. 즉 응시자 능력수준과 검사내용간 연결이 된다(Keeves, 1998; Mislevy, 1992). 이렇게 문항반응이론을 이용한 검사제작은 발견하기가 쉽지 않다. .

그런데 1차 합격자의 마지막 사람이 검사의 4개 영역들에 걸쳐서 평균 80%정도의 점수를 보이기 위해서는 문항들을 어떻게 만들어야 할까? 일반적으로 사용되는 문항반응이론에서 능력 모수에 직접 관련되는 모수는 곤란도이다. 즉 또 다른 모수인 변별도와 추측도 모수는 능력 모수와 동일한 측정체계에 있지 않다. 따라서 문항의 모수 중 곤란도 모수만을 사용하는 1모수 모형을 사용하기로 하였다. 1모수 모형은 다음 식 ①과 같다.

$$P = \frac{1}{1 + e^{-(\theta - b)}} \dots \dots \dots \textcircled{1}$$

식 ①에서 P는 “정답을 할 확률” 또는 “맞출 확률”인데, 이 값을 0.7~0.9로 하고, θ 대신에 θ_a 나 θ_b 의 값을 대입하면 문항의 곤란도 b의 범위가 산출된다. 즉 θ_a 나 θ_b 에 있는 사람이 “맞출 확률”이 70%~90%인 문항은 곤란도로서 “b”를 가지는 문항인 것이다. 이렇게 하면 주어진 능력수준(θ_a 나 θ_b)과 “맞출 확률”이 주어질 때 그에 적절한 문항의 곤란도 b를 계산할 수 있으며, 그 결과는 아래의 ②와 같다.

$$\text{문항곤란도 범위: } -1.157 \sim .433 \dots \dots \dots \textcircled{2}$$

따라서 가교집단의 능력모수에 대한 평균과 표준편차를 각각 0과 1로 할 때 바람직한 문항들의 곤란도는 -1.157과 0.433 사이에 있게 된다. 문항은 1차 제작에서 A형, B형으로, 2차 제작에서 A형, B형으로 나누어 제작하여 실시하였다. 앞으로 1차 제작 및 실시를 1차 실험평가, 2차 제작 및 실시를 2차 실험평가로 부르기로 한다.

1차 실험평가

1. 문항제작

각 영역 당 30문제씩 총 120문제를 하나의 세트로 하여 A형과 B형을 개발하였다. 언어영역에서는 기존에 고시 1차 시험 과목인 헌법과 국사를 지문으로 하는 문항들은 물론 경제, 과학, 시사 등 미리 설정된 영역에서 서두누락, 중간누락, 문단구조파악, 문단배열 등 전형적인 유형들이 사용되었다. 자료해석영역에서는 공직사회에서 다루게 되는 실용적 지문 속에서 자료의 해석이 요구되는 문항들이었다. 상황판단영역에서는 문제에 함유된 논리적 맥락 자체로서도 능력을 측정할 수 있으므로 특별히 공직사회의 맥락이 강조된 지문들을 사용하지는 않았다. 상식영역에서는 다양한 영역에 걸쳐서 기본적 지식이라고 할 수 있는 내용들이 문항화되었다.

2. 실시

2000년 7월 11일부터 7월 14일까지 다음의 세 집단에게 A형과 B형을 실시하였다. 대학생 215명, 재직자 100명, 수습사무관 220명을 대상으로 실시하였다.

3. 문항분석

실시결과에 기초한 문항의 양호도는 2차 실험평가에서 새로이 제작되어 실시되는 2개 세트의 자료수집이 끝난 후 1차 평가에서의 문항들과 2차 평가에서의 문항들을 동일한 측정체계 상에서 연결한 후에 함께 논의하기로 한다.

2차 실험평가

실험평가가 1차와 2차로 나뉜 것은 문항제작의 시간적 일정을 고려하여 가급적 실험평가를 일찍하고 정보를 활용하기 위해서였다.

1. 문항제작

또 다시 각 영역별로 A형, B형을 제작하였다. 언어, 상황판단, 자료영역에서는 30문항을 하나의 세트로, 상식영역에서는 50문항을 하나의 세트로 하였다. 상식영역이 1차 실험평가에서는 30문항이었는데 50문항으로 확대한 것은 두 가지 이유에서였다: 응답행동의 신속성, 포괄해야할 세부영역의 넓음. 1차 실험 평가시에 상식의 경우 응답행동을 관찰한 결과, 정답을 “아는가/모르는가”가 결정되면 긴 생각이 필요 없이 신속히 응답과정이 완료되는 것을 발견하였다. 즉, 상대적으로 시간이 적게 걸린다는 것이다. 그렇다고 상식문제의 성격상 언어처럼 지문을 길게 하는 것도 쉬운 일이 아니었다. 반면에 상식영역은 다른 영역에 비해서 세부영역을 보다 다양하게 설정해야 하고 또 그렇게 할 수가 있다. 따라서 응답의 신속성 그리고 세부영역 포괄의 필요성 때문에 상식문제는 2배에 가까운 50문제로 하였고, 나중에 실제로 검사시간을 측정할 때 주어진 60분간에 충분히 실시가 가능하였다.

2. 실시

전국의 공무원 준비생 366명, 재직자 54명, 수습사무관 203명이 실험평가에 참여하였다. 그리고 이후에 기술된 척도연결용 표집으로 737명의 대학생이 참여하였다.

3. 척도연결 후 결과

척도 연결 후 모든 응답자의 능력점수는 θ , 즉 표준정규분포를 따르는 모수의 측정치로 표시된다. 이 값은 대체로 -3과 +3사이에 있게 된다. 그러면 집단별 능력모수 측정치를 검사 영역별로 보기로 한다.

(1) 언어능력

	평균	표준편차	인원
대학생	0.148	0.668	507
재직자	0.323	0.662	120
수습사무관	0.383	0.657	370
전체	0.256	0.672	997

전체평균은 0.256이었고 모든 집단이 가교집단보다 높은 언어능력점수를 보였다. 또한 대학생보다는 재직자나 수습사무관들이 높은 능력점수를 보인다.

(2) 상황판단력

	평균	표준편차	인원
대학생	0.004	0.766	482
재직자	0.010	0.690	127
수습사무관	-0.236	0.876	378
전체	-0.087	0.809	987

전체평균은 가교집단보다 약간 낮았다. 특히 수습사무관들에게서 낮은 상황판단력점수가 관찰되었는데, 이것은 그들이 시험 준비기간 중 익숙해온 문제들과 상황판단영역의 문제들이 내용과 유형에서 서로 다름을 의미할 수 있다. 또는 시험준비로 인하여 상황판단력의 활용 또는 활성화가 억제되어 있다고 볼 수도 있다.

(3) 자료해석력

	평균	표준편차	인원
대학생	0.138	0.782	470
재직자	0.351	0.731	115
수습사무관	0.093	1.012	370
전체	0.146	0.876	955

전체평균이 0.146이었고 모든 집단이 가교집단보다 높은 자료해석능력을 보였다. 특히 재직자가 훨씬 높았음은 지문들이 실제 업무에서의 분위기가 실려있는 것들이었음을 시사한다.

(4) 상식수준

	평균	표준편차	인원
대학생	0.126	0.564	478
재직자	0.719	0.500	132
수습사무관	0.504	0.645	381
전체	0.350	0.631	991

전체평균은 0.350으로서 가교집단보다 아주 높았다. 특히 재직자와 수습사무관들의 상식수준은 대학생에 비해서 거의 1표준편차만큼 높았다.

개발 결과

각 영역별로 종합 분석한 결과를 논의하기로 한다.

1. 언어영역

인간의 언어처리과정에서 한 문장을 처리하기 위해서는 먼저 각 단어의 의미를 파악하고, 그 단어들을 문법 법칙에 의하여 결합하여 최종적으로 의미를 파악하여야 한다. 여기에 관여하는 인간의 언어능력으로 어휘력, 문법적 지식, 문장에 근거한 추론 능력 등이 있다. 그리고

이렇게 이해된 문장들은 논리적 관계에 의해 결합되고 통합되어 최종적으로 텍스트(text)에 대한 의미를 파악하게 된다.

이 과정에는 주어진 텍스트에 관한 사전 지식, 논리에 의한 의미의 파악 능력, 요약과 추론 능력 등이 관여되게 된다. 따라서 인간의 언어처리능력은 기억력, 어휘력, 문법적 지식, 추론 논리 능력 등을 반영하는 복잡한 개념적인 구성체이다. PSAT의 언어영역에서는 어휘력이나 문법적 지식과 같은 문장 수준의 처리 능력을 다루기보다는 텍스트의 처리와 관련된 능력을 측정하고자 하였다.

우선 문항분석 결과를 요약하기로 한다. 검사는 1차와 2차로 나누어 실시하였으므로, 그에 따라 요약한다. 1차 검사는 A형과 B형으로 30문항씩 나누어 서로 다른 응시자들에게 실시하였으며, 각 문항에 대한 이들의 반응을 ITEMAN 3.5 와 BILOG (Mislevy & Bock, 1986)를 이용하여 문항분석 하였다. 문항분석결과 정답률은 0.14에서 0.99사이의 값을 보이고 있었으며, A형과 B형의 평균 정답률은 각각 0.66과 0.60이었다. 이러한 결과는 A형과 B형의 각 문항에 대한 평균 정답자 비율이 전체 응시자들의 66%와 60%임을 의미하는 것으로서, 이러한 정답률은 문항의 난이도가 중간 정도임을 보여주는 결과이다.

2차 검사 또한 1차 검사에서와 같이 A형과 B형으로 30문항씩 나누어 서로 다른 응시자들에게 실시하였으며 각 검사의 1번에서 6번까지의 문항들은 척도의 연결을 위하여 1차 검사에서 사용하였던 문항들을 그대로 이용하였다. 새로운 문항들에 대한 반응을 ITEMAN 3.5와 BILOG 를 이용하여 문항분석 하였다.

문항분석 결과 A형과 B형의 평균 정답률은 각각 0.67과 0.73으로서, A형과 B형의 각 문항에 대한 평균 정답자 비율이 전체 응시자들의 67%와 73%임을 의미하는 것이다. 이러한 정답률은 문항의 난이도가 1차 검사에 비해 쉬웠음을 보여주는 결과로서, 1차 검사 이후 문항의 난이도 수준을 다소 낮추었기 때문이다.

제출용 문항 선정은 1차 검사와 2차 검사에 대한 개별적 분석 과정에서 선정된 문항들을 토대로 실시하였다. 즉 1단계에서 정답률과 난이도가 극단인 45개 문항을 제외하였고 2단계에서는 검사의 문항이 남성과 여성 중 어느 한 쪽 성에 대해 차별적으로 유리하게 작용하는지를 알아보기 위하여 차별기능문항에 대한 분석을 실시하였으며, 그 결과 여성에게 불리한 4개 문항들은 추가로 제외하였다. 이러한 분석결과를 토대로 앞에서 제시한 문항선정 기준에 따라 A급과 B급의 최종문항을 선정하였다. 결과로서 A급이 34개 문항, B급이 26개 문항이었다.

A급 문항 34개중 30개를 제출용 A세트로, 나머지 4개와 B급 문항 26개를 제출용 B세트로 하였고, 나머지 C세트는 새로이 개발하였다. 물론 C세트는 현장실시가 되지 않은 부분이다. 각 세트별 구성은 다음과 같다. 가급적 A세트의 내용분야가 평행하도록 하였으며 유형은 자유롭게 다양화하였다. 제작되어 선정된 문항은 A세트 30문항, B세트 30문항, C세트 30문항이 있다. 이러한 문항제작에 반영된 내용분야와 유형을 정리하면 아래와 같다.

언어영역에서의 내용과 유형

1) 내용분야

경제, 고전문학, 과학, 교육, 국사, 국제, 사회, 사회학, 서양사, 스포츠, 인류학, 정치, 통일, 행정, 행정법, 행정학, 헌법, 현대문학, 환경(19개 분야)

2) 유형

글의 이해, 다음 주제, 무관련 문단찾기, 문단구조 파악, 문단배열, 반론, 배열후주제, 비판, 앞문단누락, 전제추론, 제목찾기, 주제찾기, 중간단락 누락, 추론되는 내용, 후속문단 (15개 유형)

2. 상황판단영역

상황 판단력은 연역추리력, 문제해결, 판단 및 의사결정 능력을 측정하는 문제로 구분되어 있다. 연역추리력은 주어진 사실(전제)들로부터 논리적으로 정당한 결론을 도출해 낼 수 있는 능력이다. 이 연역추리는 정보를 주어진 형태로 처리하는 것이 아니라, 주어진 정보에 근거하여 보다 풍부한 새로운 정보를 산출하게 하는 것으로서, 정보 처리 과정의 필수적인 요소이다.

간단한 예를 들면, “A는 B보다 크다”와 “B는 C보다 크다”라는 정보가 주어지면 인간은 이 정보에 근거하여 “A는 C보다 크다”라는 새로운 정보를 산출하게 된다. 이 연역추리는 대부분의 사람들이 자동적으로 행할 수 있지만, 복잡한 논리구조를 가진 정보에서는 연역추리가 용이하지 않기 때문에, 훈련을 필요로 한다.

판단과 의사결정은 주어진 정보와 주어진 정보에서 유도된 정보들을 정확하게 판단하고, 그 판단에 근거하여 가장 합리적인 의사결정을 하는 과정이다. 문제풀이의 과정은, 머리 속에서 주어진 문제를 표상하고 주어진 초기 상태와 도달하기 원하는 최종 목표상태를 명확히 정의하여 초기상태에서 목표상태로 이전시키는 조작자 찾기의 과정으로 구성되어 있다. 따라서 문제해결능력은 문제에 적절한 표상을 형성하고 목표상태에 도달하게 하는 적절한 조작자를 찾아내는 능력을 포함하고 있다. 구체적인 내용들은 다음의 표에 제시되어 있다.

내 용	유 형
논리추리	논리구조
문제해결	논리적 타당성
의사결정	평가
판단	논증
유추	결론유도
	게임이론
	기획력
	해석
	해결

기본적인 분석 절차는 언어영역의 절차를 따랐다. 1차 검사의 문항분석 결과 각 문항과 척도점수간 점이연(Point-biserial) 상관계수 평균을 보면, A형에서 0.328, B형에서 0.277로 무난한 편이다. 응답률의 평균은 A형 62.5%, B형 57.2%로서 검사가 쉬운 편은 아니었음을 나타낸다. 2차 검사의 문항분석 결과 각 문항과 척도점수간 점이연 상관계수 평균은 A형에서 0.332, B형에서 0.370이었고 응답률의 평균은 A형 63.4%, B형 65.0%로서 1차 검사보다는 난이도가 다소 낮아졌다는 것을 알 수 있다.

제출용 문항선정은 언어영역에서 기술한 바와 동일하게 진행되었다. 이미 제시된 문항선정 기준에 따라 A급(31개 문항)과 B급(29개 문항)의 최종문항이 선정되었다. 최종적으로 선정된 문항들의 내용 특성과 통계적인 문항 특성을 고려하여 이들 문항과 평행한 제 3의 문항세트(C세트)가 제작되었다. 이 문항들은 최종 선정된 문항들과는 동일한 능력을 측정하고 유사한 변별도와 난이도를 유지할 수 있도록 구성하였다.

3. 자료해석영역

현대 사회에서는 정보가 언어라는 형식으로 제시되기도 하지만 많은 경우에 수치나 그래프와 같은 형식으로 제시된다. 보통 하나의 그래프나 표에는 많은 정보가 함축적으로 제시되기 때문에, 이렇게 제시되는 정보를 정확하게 이해하기 위해서는 계산이나 추론을 통해 직접적으로 제시되지 않는 정보를 추출할 수 있어야 한다. 특히, 공직자들은 부처를 막론하고 다양한 자료들을 자주 접하게 되므로, 이들 자료로부터 필요한 정보를 추출할 수 있는 능력의 중요성은 매우 크다 하겠다.

따라서, 자료해석력에서는 공직자가 업무 속에서 접할 수 있는 문화, 경제, 사회, 정치, 법률, 환경 등 다양한 분야에서의 자료들을 제시하고, 이들 자료를 정확하게 이해하기만 하면 복잡한 계산 없이도 문제를 풀 수 있도록 함으로써, 단순한 계산능력이 아닌 자료에 대한 종합적인 분석능력과 추론능력을 측정하고자 하였다.

문제의 유형은 크게 2가지로 구분될 수 있다. 첫 번째는 제시된 진술문들의 진위를 파악하는 유형으로서, 세부적으로는 3개 또는 4개의 진술문들 중 옳거나 또는 그른 것들을 모두 찾는 유형과, 5개의 보기 중 하나의 정답을 찾는 유형이 있다. 이러한 문제 유형은 각각의 진술문들의 진위판단이 실제로는 하나의 문항역할을 수행하는 것이라 할 수 있으며, 제시된 진술문이 주어진 자료를 통해 유출될 수 있는 진술인지를 판단하는 연역적 추론과정이 요구된다는 점에서 선지해석 중심형 문항이라 할 수 있다. 두 번째는 자료를 통해 문제가 요구하는 정보를 추론하거나 약간의 계산과정을 통해 정답을 찾아내는 유형으로서, 이 유형은 전자의 유형과 달리 자료에 대한 분석을 통해 해답을 도출해내는 귀납적 분석과정을 요구한다는 점에서 본문해석 중심형 문항이라 할 수 있다.

분석방법은 언어/상황판단 영역과 동일하였다. 1차 검사에서 문항분석 결과 정답률은 대부분 .20에서 .70 사이의 값을 보이고 있었으며, A형과 B형의 평균 정답률은 각각 .39와 .42였다. 이러한 결과는 A형과 B형의 각 문항에 대한 평균 정답자 비율이 전체 응시자들의 39%와 42%임을 의미하는 것으로서, 이러한 정답률은 문항의 난이도가 적절함을 보여주는 결과이다. 또한, 변별력을 나타내는 점이면 상관 분석 결과, 대부분의 상관값들이 .20과 .40 사이에 위치하고 있었으며, A형과 B형 검사의 평균 점이면 상관관계수는 각각 .23과 .27의 값을 보이고 있었다. 이러한 수치는 이 검사의 총점과 각 문항간의 상관이 그리 높지 않음을 의미하는 것이다. 그러나, 이러한 결과를 보인 이유는 몇몇 문항들이 응시자들의 자료해석력을 측정하는데 제대로 기능하고 있지 않았기 때문으로, 실제로 1차 검사 문항들 중 최종문항 선정과정에 포함될 문항들만을 추출하여 다시 계산한 평균 점이면 상관관계수는 A형과 B형에서 각각 .385와 .410로 비교적 높았다.

2차 검사 문항분석 결과 정답률은 대부분 .30에서 .80 사이의 값을 보이고 있었으며, A형과 B형의 평균 정답률은 각각 .49와 .54로서, 이 결과는 A형과 B형의 각 문항에 대한 평균 정답자 비율이 전체 응시자들의 49%와 54%임을 의미하는 것이다. 이러한 정답률은 문항의 난이도가 1차 검사에 비해 다소 쉬웠음을 보여주는 결과로서, 1차 검사 이후 분석 결과를 토대로 응시자들의 정답률(난이도)을 .50의 수준까지 높이고자, 문항의 난이도 수준을 다소 낮추었기 때문이다.

점이면 상관 분석 결과, 대부분의 상관값들이 .25와 .50 사이에 위치하고 있었으며, A형과 B형 검사의 평균 상관관계수 각각 .34와 .35였다. 이러한 수치는 이 검사의 총점과 각 문항간의 상관이 양호하다는 것을 의미하는 것으로서, 대부분의 문항들이 응시자들의 자료해석력 측정을 위해 적절히 기능하고 있음을 의미한다. 더욱이, 최종문항 선정과정에 포함될 문항들만을 추출하여 다시 점이면 상관을 분석한 결과에서는 1차 검사결과에서와 마찬가지로 A형과 B형이 각각 .391과 .417의 비교적 높은 상관을 보이고 있었다. 언어/상황판단 영역에서 기술된 바와 동일한 방식으로 C세트의 문제가 개발되었다.

4. 상식영역

사회 상식 영역은 5급 공무원으로서 알아야 할 다양한 분야의 기본적인 지식을 측정하는 검사로서, 단편적이고 피상적인 지식이 아니라 사회적으로 관심을 받는 개념·현상들에 대하여 정확하고 깊이 있는 이해를 하고 있는지를 측정하고 있다. 사회적으로 이슈가 되고 있는 개념·현상들에 대한 이해를 물음으로써 이 영역은 1차적으로 지식에 대한 평가를 하지만 부가적으로 지원자가 가지고 있는 사회에 대한 관심 정도를 측정할 수 있다.

행정/기술직의 각 직제별로 출제영역을 골고루 배분하고자 하였으며, 이를 위해 정부 조직의 17개 부처(외교통상부, 재정경제부, 법무부, 통일부, 국방부, 행정자치부, 교육부, 과학기술부, 문화관광부, 농림부, 해양수산부, 산업자원부, 정보통신부, 환경부, 보건복지부, 노동부, 건설교통부)와 관련성이 있는 문제가 고르게 출제되도록 하였다.

출제의 공정성 및 편파성 시비를 방지하기 위해 신문이나 언론 매체에서 언급된 내용을 출제 자료로 활용하였다. 문항의 출제는 해당 출제영역의 기본적인 지식에 해당하는 일반적인 문제와 최근 1년 이내에 사회적으로 관심을 끌었던 시사성이 높은 문제의 두 가지 형태로 하였으며, 전체적인 비중은 지원자의 사회적인 관심도를 측정하는데 강조를 두기 위하여 시사성이 높은 문제가 70% 정도가 되도록 하였다. 물론 시사성이 높은 문제라고 하더라도 기본적인 지식이 포함되므로 엄격하게 두 가지 문제유형을 구분하기는 쉽지 않다.

사회상식 문제의 특성상 다양한 분야의 지식을 하나의 검사에 포함하게 되므로 다른 검사 유형들과는 달리 사회상식 문제에 대해서는 문항반응이론보다는 고전 검사이론에 의한 분석 및 성차별 문항 분석을 통해 문항을 선정하는 것이 타당한 것으로 판단되었다.

그러나, 다른 검사 영역과 동일한 분석과정을 적용해 보고 그 결과를 검토해 보기 위하여 문항반응이론을 이용한 BILOG-MG 프로그램을 통해서 피험자 능력집단에 독립적인 개별문항의 곤란도를 산출하여 A급과 B급을 선정하였다. 그 결과, A급 문항은 27문항이 선정되었으며, B급 문항은 15문항이 선정되었다. 이는 전체 130개의 문항을 개발하여 42개의 문항(32.3%)만 양호한 문항으로 선별해 낸 결과로, 타 검사 영역에서 문항반응 이론을 적용하여 다수의 최종 문항을 선정해 낼 수 있었던 결과와 대비된다.

개발된 검사의 신뢰도와 타당도

1. 신뢰도

최종적으로 선택된 문항들만을 대상으로 신뢰도가 분석되었다. 특히 PSAT는 기본적으로 준거 참조적으로 사용되므로 분류의 일치도가 신뢰도로 사용되었다.

(1) 준거참조적 신뢰도 분석

준거 참조적 신뢰도 분석에서는 외적 준거에의 일치도가 곧 신뢰도이다. 우선 PSAT에서는 지원자의 10%~15%가 80%이상의 점수를 받은 것을 검사의 양호도에 대한 준거로 제시하고 있다. 우리는 각 영역별로 4개 자료 세트(1차-A, 1차-B, 2차-A, 2차-B)에서 세트내 총점 기준으로 80%이상 맞추는 사람의 비율을 다음과 같이 구하였다.

표 6. 80% 이상 맞은 사람의 비율(%)

	1차-A	1차-B	2차-A	2차-B	영역별 중앙값	전체 중앙값
언어	11.48 (N=183)	18.01 (N=161)	10.86 (N=221)	13.22 (N=227)	12.35	
상황판단	6.42 (N=187)	4.76 (N=168)	11.11 (N=216)	26.70 (N=236)	8.77	
자료해석	3.87 (N=181)	8.28 (N=157)	10.55 (N=218)	16.97 (N=218)	9.42	10.89
상식	22.16 (N=194)	24.57 (N=175)	37.91 (N=211)	33.33 (N=216)	28.95	

표 6에서 보면 80% 이상 맞추는 사람들의 비율이 언어 영역에서 10%~15% 이내에 있고, 상황판단과 자료해석영역에서는 10% 미만이 되며, 상식 영역에서는 25%를 넘는다. 즉, 준거에 비추어 볼 때 상황판단과 자료해석영역의 문항들이 상대적으로 좀 어려운 편이고 상식영역이 좀 쉬운 편이다. 그런데 이 자료는 단지 실험평가에 지원한 사람들을 대상으로 한 것이며 모든 지원자가 4개 영역에 모두 다 응답한 것도 아니므로 4개 영역의 총점을 가지고 80%이상이 몇 명이 나오는지 계산할 수 없었다. 그러나 이 비율에 대해서 각 영역별 중앙값에 기초해서 전체적 중앙값을 구하면 10.89%로서 10%~15% 범위에서 벗어나지는 않는다.

다음에 각 영역에서 응시자들을 80% 이상을 받은 사람의 집단과 그 미만을 받은 사람의 집단으로 분류하여, 개인에 대한 분류에서 A형 검사와 B형 검사간에 일치하는 정도를 측정 한 결과를 제시한다. 이 결과는 PSAT의 준거참조적으로 사용할 때, 또 하나의 신뢰도가 된다. 1차 검사에서는 A형 검사와 B형 검사를 모두 응답한 사람들이 없었고, 2차 검사에서는

영역별로 약 50여명씩이 있었으므로 이들의 자료에 기초해서 분류비율의 일치도를 구하면 다음과 같다.

표 7. 분류비율의 일치도

	언어	상황판단	자료해석	상식
N	59명	45명	52명	57명
비율	84%	60%	92%	81%

일반적으로 일치도의 값이 85%일 때 높다고 하는 것을 참고로 하면 언어 영역과 자료해석 영역에서의 일치도는 높은 편이고, 상식 영역도 높은 편에 가깝다. 단지 상황판단 영역만은 높지 않다.

2. 타당도

(1) 구성개념의 수렴과 변별

문항개발이라는 이 연구의 목적상 구성개념타당도를 중심으로 하였고 준거타당도는 후속 연구의 몫이 될 것이다.

4개 영역의 변별성: 4개 영역을 네 시간에 걸쳐서 모두 응답한 사람만을 가지고 4개 검사 간 상관을 구하면 다음과 같다.

표 8. 검사영역간 상관의 평균

영역(문항수효)	언어	상황판단	자료해석	상식
언어(60)	1.00			
상황판단(60)	.24	1.00		
자료해석(60)	.17	.18	1.00	
상식(42)	.21	.16	.14	1.00


표 14에서 보면 언어능력은 상황판단 및 상식과 0.20을 넘는 크기의 상관이 있고 자료해석은 언어능력과 0.17, 상황판단과 0.18, 상식과 0.14의 상관이 있다. 이들 크기는 이 4개 영역간 관련성을 보여주긴 하지만 관계의 정도가 크다고 할 수는 없다. 따라서 4개 영역간에는 충분한 변별성이 있다.

(2) 사무관들의 행정고시/기술고시와의 관계

현행 고시 제도와 PSAT의 관련성을 탐구하기 위하여 PSAT의 각 영역의 점수와 현행 고시의 1차 점수의 상관계수를 구하였다(표 9 참조). 고시의 1차 점수와 PSAT의 관계를 살펴보면, 언어영역과 상식영역에서 의미있는 상관을 보이며, 상황판단력 영역과도 상관의 경향성이 보인다. 이러한 상관분석의 결과는 현행 고시제도가 PSAT의 영역 중 언어영역과 상식영역을 반영하고 있음을 알 수 있다. 현행 고시의 과목 구성으로 볼 때 이러한 결과는 충분히 예측할 수 있는 결과이다. PSAT가 기존의 고시제도와 차별화되는 영역으로 상황판단력과 자료해석력을 들 수 있고, 직무분석 결과에서 이 영역이 공직자로서의 업무 수행에 필수적으로 밝혀졌음을 감안할 때 PSAT가 지니는 장점으로도 해석될 수 있다.

표 9. 행정고시/기술고시 1차점수와와의 상관

		1차예비검사 A유형과 1차 고시점수와의 상관	1차예비검사 B유형과 1차고시점수와의 상관	2차예비검사 A유형과 1차 고시점수와의 상관	2차예비검사 B유형과 1차 고시점수와의 상관
언어 영역	상관계수	0.090	0.161	0.035	0.008
	P- value	0.300	0.091	0.743	0.008
	N	136	111	93	97
자료해석력	상관계수	-0.009	0.025	0.161	0.004
	P- value	0.921	0.797	0.117	0.967
	N	136	111	96	90
상황판단력	상관계수	0.029	0.096	0.083	0.189
	P- value	0.733	0.318	0.429	0.058
	N	136	111	93	101
상식	상관계수	0.102	0.297	0.001	0.079
	P- value	0.236	0.002	0.001	0.079
	N	136	111	98	94

 : 실질적으로 유의하게 해석될 수 있는 크기의 상관 계수들 (p<.05)

결론 및 논의

본 연구의 목적은 공직 적격성을 측정할 수 있는 검사(PSAT)를 개발하는 것이었다. 바람직한 공직자를 급속히 발달하는 과학기술이나 지식을 학습할 수 있는 능력을 지니고 있고 급격하게 변하는 사회제도가 의식에 적응할 수 있는 능력을 갖춘 사람으로 정의하고, 이러한 능력을 측정하는 것을 공직 적격성 검사의 목표로 삼았다.

이론적 분석과 함께 우리 나라와 외국의 사례를 조사하여 공직 적격성 영역으로 언어, 자료해석, 상황 판단, 상식의 네 영역이 평가영역으로 산출되었고, 이 평가영역은 검사 개발을 위한 직무조사 결과와 일치되었다. 다만 상식의 영역은 여전히 지식의 정도를 측정하고 있기

때문에 기존의 시험과 차별성이 없다는 지적이 있을 수 있으나 상식이 공직자의 최소한의 기본지식을 측정하는 척도로 사용할 수 있다는 판단 하에 상식영역을 공직적격성의 한 영역으로 포함시켰다.

그러나 상식의 문제는 변별성에 강조를 두기보다는 문제의 곤란도를 낮게 하고 다양한 영역에 대한 질문을 함으로써 사회제도나 사회현상에 관심이 있는 사람이면 통과할 수 있는 문제를 제작하려고 하였다. 즉 변별보다는 예비심사(screening)에 중점을 두는 문제를 제작하였다. 따라서 이러한 상식의 영역은 지식과 더불어 사회에 대한 관심의 정도를 측정하는 척도로도 사용될 수 있을 것이다.

이렇게 추출된 네 영역을 측정할 수 있는 문항들이 심리측정이론에 근거한 과학적 방법론을 도입하여 개발되었고 각 문항의 타당성이 검증되었다. 타당성 검증을 위하여 대학생, 수습사무원 그리고 공직에 재직하는 사람들을 대상으로 1차, 2차의 예비 검사를 실행하였고, 이 예비 검사의 결과를 바탕으로 하여 최종적인 문항들이 선택되었다. 이러한 문항들은 신뢰도와 타당도에서 과학적 방법론에 근거한 검사로서 수용할 만한 수준이었다. 특히 타당도와 관련하여, 대학생과 수습사무원, 그리고 재직자의 점수를 비교하였는데, 언어능력, 자료해석능력, 그리고 상식의 영역에서 재직자, 수습사무원, 대학生の 순서로 높은 점수를 얻었다. 재직자의 점수가 수습사무원이나 대학生の 점수보다 높다는 결과는 공직에 재직하는 사람이 그렇지 않은 사람보다 공직적격성을 더 많이 지니고 있음을 의미하므로, PSAT가 공직적격성을 측정하고 있다는 간접적인 증거가 될 수 있다. 그러나 논리력, 추리력, 그리고 문제해결능력을 측정하는 상황 판단력 영역에서는 재직자와 대학생들 간에는 차이가 없었다.

이 결과는 상황판단력이 기존의 고시가 측정하지 못하는 것을 측정하고 있으며, 이에 대한 면밀한 검토가 필요함을 시사한다. PSAT가 현행 고시의 1차 시험을 대체하려는 목적으로 제작되었으므로 현행 1차 고시와의 관계성도 중요하다. 이에 대한 분석에서, 고시의 1차 점수는 PSAT의 언어와 상식 영역에 유의한 상관을 보였고, 자료해석력과 상황판단력에는 낮은 상관을 보였다. 따라서 자료해석력과 상황판단력은 기존의 1차 시험에서 측정되지 않은 능력으로 해석되었다.

이러한 연구절차와 결과에 근거하여 PSAT의 특징을 나열하면 다음과 같다.

첫째, PSAT는 과학적 연구 절차에 의해 개발된 검사이다. 심리측정이론의 절차에 따라 문항개발과 문항 분석을 하고 타당도 연구를 걸쳐서 검사가 개발되었다.

둘째, PSAT는 단순한 암기 지향적 문제를 지양하고 지식사회에 신속히 적응할 수 있는 학습능력과 문제 해결 능력을 측정하는 데 중점을 둔 검사이다. 따라서 현재와 같이 고시공부에만 집중하는 것은 PSAT에서 우수한 성적을 받는데 그리 큰 도움이 되지 못하며, 여러 방면의 지식을 쌓고 논리적 사고력을 키우며 학습능력을 늘리는 것이 PSAT에서 높은 성적을 얻는데 도움을 줄 것이다.

PSAT의 이러한 특성은 다음과 같은 순기능을 가진다.

첫째, 암기위주의 고시제도로 인한 대학 교육의 파행성을 해결하는데 도움을 줄 수 있다. 둘째, 충실한 교육을 받고 교양 내지는 전문분야의 지식을 쌓은 사람들이 공직에 지원할 경우 별도의 부가적인 준비 없이 PSAT에 응시할 수 있어서, 많은 인재들이 고시에 응시하는 것을 유도할 수 있다. 셋째, 고시에 실패하거나 준비 도중에 포기하는 경우에도 다른 직종으로의 전환이 용이하여 고시공부에 대한 기회비용을 감소시킬 뿐만 아니라 고시공부로 인한 사회적 비용을 절감할 수 있을 것이다. 이러한 모든 측면은 우수한 인재들이 공무원 시험에 응시하게 하는 유인가를 제공함으로써 보다 많은 인재를 공직에 확보하는데 도움을 줄 것이다.

참고문헌

- 중앙인사위원회 (2000). *공직 적격성 테스트 문제개발 용역제안 요청서*. 서울: 저자.
- Baker F. B. (1985). *The basics of item response theory*. Portsmouth, NH: Heinemann.
- Gardner (1983). *Frames of minds*. New York: Basic.
- Keeves, J. P. (1998). Scaling achievement test scores. In Walberg & Haertel (Eds.), *The International Encyclopedia of Educational Education*, 308-318. Pergamon Press.
- Mislevy, R. J. (1992). *Linking educational assessment: Concept, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.
- Mislevy, R. J. & Bock, R. D. (1986). *PC-BILOG: User's Guide*. Mooresville, IN: Scientific Software.
- Schmitt, N., & Chan, D. (1998). *Personnel selection: A theoretical approach*. Beverly Hills, CA : Sage Publications.
- Sternberg, R. J. (1995). Theory and measurement of tacit knowledge as a part of practical intelligence. *Zeitschrift für Psychology*. 203, 319-334.
- Wilson, D. T., Wood, R., & Gibbons, R. (1991). *Testfact: Test scoring, item statistics, and item factor analysis*. Chicago, IL : Scientific Software International Inc.