

**OECD 국제성인역량조사(PIAAC) 관련
데이터 관리 교육 및 IT-C 회의 참가 결과 보고**

2010. 03.

목 차

I. 출장개요	1
II. 출장결과	2
1. 데이터 관리 교육/훈련	2
가. 데이터 관리 업무	2
나. 데이터베이스 구조	4
다. 데이터 레코드	4
라. DME software	6
2. IT-C 워크숍	7
가. 추진 경과	7
나. 결과	7
III. 향후 추진 일정	8

OECD 국제성인역량조사(PIAAC) 관련 데이터 관리 교육 및 3차 IT-C 회의 참가 결과 보고

I. 출장개요

출장목적

- OECD가 주관하는 성인능력 국제비교 프로젝트(Programme for the International Assessment of Adult Competencies, PIAAC)의 통계대행과 관련하여 2010년 예비조사의 자료처리를 위한 데이터 관리 교육/훈련(Data Management Training) 및 3차 IT-Coordinator 회의에 참석

참가회의 : PIAAC Data Management Training PIAAC 3rd IT Coordinators Meeting

회의주관 : OECD PIAAC 컨소시엄

출장 수행내용

- PIAAC 2010년 예비조사 데이터 처리를 위한 교육/훈련 참석
- PIAAC 조사 시스템 통합 및 데이터 전송을 위한 워크숍 참석
- 예비조사를 위한 시스템 최종 점검 및 IT 지원 관련 회의 참석

출 장 자 : 2명

- 통 계 청 : 통계대행과 한혜은주무관
- 한국직업능력개발원 : 임 언 박사

출장기간 : 2010년 2월 20일 ~ 2월 27일 (6박 8일)

출장지역 : 독일 (함부르크)

II. 출장결과

1. 데이터 관리 교육/훈련

가. 데이터 관리 업무

○ 컨소시엄의 역할

- IEA(The International Associational Achievement) Data Processing and Research Center(DPC)는 인터내셔널 레벨의 자료 처리와 참가국의 데이터 관리 담당
 - PIAAC 데이터 처리를 위한 국제 표준 지침 및 권고 제안
 - 소프트웨어, 코드북, 매뉴얼 제공
 - 예비조사 및 본 조사의 데이터 처리를 위한 교육
 - 참가국의 데이터 처리 지원
 - 참가국의 Data Management 품질 관리
 - 인터내셔널 레벨의 데이터 관리, 처리, 클리닝
 - 데이터베이스 보급 및 분석 준비

○ 참가국의 역할

- NPM의 책임
 - 국제 표준 지침에 따라 품질이 보증된 데이터베이스 생산
 - 데이터 관련 업무는 높은 전문성을 요구하므로 NPM이 NDM (National Data Manager)을 지정하는 것을 권고
- NDM의 역할
 - 참가국의 데이터 처리 업무 총괄
 - 데이터 전송 이후에 IEA DPC와 직접 연락하며 컨소시엄 업무 지원
- NDM의 자격
 - 데이터베이스에 대한 전문적인 기술과 다양한 포맷의 파일을 전환하고 처리할 수 있는 능력
 - PIAAC 시스템, 조사도구, 평가 설계 및 조사 연구에 대한 지식과 기술
 - 대규모 조사의 데이터 관리 계획, 수행 및 감독 경험
 - 본 조사를 포함한 프로젝트 전체 기간 동안 동일한 사람이 수행할 수 있도록 안정된 근무 조건 요구

- NPM과 같은 조직에서 근무하는 것을 권고
 - 데이터 관련 커뮤니케이션을 위해 영어 능숙
 - PIAAC SharePoint의 Contacts list에 NDM 정보 등록
 - NDM은 매뉴얼 자료 입력과 조사 기간 동안 데이터 관련 업무를 담당할 수 있는 직원의 지원이 필요하며, 담당자에게 데이터 관리를 위한 교육 실시
- 데이터 관리에 필요한 자료
- 주요 자료(Key Resources)
 - Data Management Manual : 데이터 관리 업무에 필요한 매뉴얼
 - Data Management Expert(DME) software : 참가국의 데이터를 처리하기 위한 데이터 관리 소프트웨어
 - Codebook : 모든 조사도구(배경설문조사 문항, 컴퓨터평가 문항, 서면평가 문항) 및 work-flow와 관련된 변수들에 대한 International English source 버전의 데이터베이스 정의
 - 서면 평가 검사지(Paper instruments), 응답 내용 기록지(capture sheet), 채점지 (scoring sheet) : 서면 평가 결과 및 채점 결과 입력 시 활용
 - 국가별 데이터 처리를 위한 자료
 - BQAS(Background Questionnaire Adaptation Sheet) : 참가국에서 변경하거나 추가한 배경설문 문항에 대해 정의한 문서로 National codebook 작성을 위해 필요
 - Conversion : 인터내셔널 변수로 변환해야 하는 변수에 대한 정의
 - 기타 자료
 - PIAAC-TAO Data Import and Export : TAO에서 사용한 변수의 정의 및 포맷에 대해 상세히 설명하는 문서로써, 조사결과를 DME software를 이용하여 데이터베이스로 가져오기 할 때 활용
 - external Coding Process 관련 문서 : Coding이 필요한 변수의 내보내기, Coding data를 DME software를 이용하여 데이터베이스로 가져오기 등
- 국제 기준 및 지침
- Data Capture, Management, Editing, submission, confidentiality and security에 대한 지침은 국제 기술 표준 및 지침(TSG)의 11~13장에 수록
 - TSG를 준수하지 못하는 경우에는 권소시업에 사전 보고

나. 데이터베이스 구조

○ 데이터 유형(Data Sources)

- 표본 설계 정보(SCF)
- 대상 초기화 정보
- 배경 설문 조사 결과
- ICT 모듈에 대한 응답자 활동
- 컴퓨터 기반 평가에서 응답 내용, 행동, 자동 채점 결과
- 무작위 할당, 현황 코드, 시간 정보 등 workflow 정보
- 선택된 응답, 항목별 소요 시간, 배경설문조사에서의 언어 변경, 탐색, 오류 수정 알림 등 행동과 관련된 정보
- 서면 평가 및 기초 읽기 능력에서 응답자의 답변
- 서면 평가 및 기초 읽기 능력에서 응답 결과 및 채점 데이터
- 교육, 직업, 산업과 관련된 응답의 코딩 데이터

- 데이터 특징

- 조사는 다양한 경로로 진행
- 일부 데이터는 조사 진행 중에 생성되고, 일부는 조사 후에 생성
- PIAAC 데이터베이스는 Multiple tables로 구성된 single database로써, 각각의 테이블은 조사의 하나 이상의 구성요소와 일치
- 테이블이 생성된 후에는 분석을 위한 “flat” 파일로 결합

다. 데이터 레코드(Data records)

○ 개요

- 데이터 레코드는 데이터 셋의 행과 일치하며, 하나 이상의 유일한 식별자에 의해 구분됨
- 응답자에 따라 설문 경로가 다르며, 한 명의 응답자에 대해 여러 개의 데이터 셋이 생성되지만, 반드시 모든 데이터 셋에 포함되지는 않음
- 각각의 변수들은 응답자의 설문 경로에 따라 하나의 데이터 셋에 유효값 또는 결측값으로 저장

○ 유효값(Valid values)

- PIAAC에서 수집한 정보에 대한 개별값으로 변수 정의에 따라 데이터베이스에 저장

- PIAAC은 대부분의 변수에 대해 변수 정의(schemes) 부여

○ 결측값(Missing values)

- 설계에 의한 결측값

- 조사 설계에 의해 한 명의 응답자가 모든 항목에 응답하지 않아 생성
- 국가별로 표본설계 방법이 달라 생성(예, 개인 추출 시 가구 ID는 결측)
- 평가 유형에 따라 응답하는 항목이 다름(C01~21, PP1~4)
- 배경설문조사 소요시간이 길어 4가지 문항군을 무작위로 응답하도록 설계

- 응답에 의한 결측값

- 표본으로 선정된 가구/개인이 거절한 경우 - 표본 정보만 존재
- 응답자가 조사 도중 중단한 경우 - 조사의 일부분에 대한 정보만 존재
- 모든 조사를 완료했으나, 항목에 따라 '모름' 이나 '거절'로 응답한 경우
- 응답자의 능력이나 동의 여부에 따라 평가 방법(컴퓨터/서면)이 다른 경우

- 자료 입력이나 데이터 처리 과정에서 발생한 항목 결측값

- PIAAC 조사 시스템의 “가져오기”나 데이터 처리 시 오류 발생
- 조사결과를 시스템에서 “가져오기” 하였으나, 내셔널 데이터베이스로 통합 되지 않은 경우
- 내셔널에서 인터내셔널로 전환이 안 되거나, 전환에 문제가 있는 경우
- 코딩, 표본 설계 데이터처럼 조사 후에 생성되는 변수에 문제가 있는 경우
- 참가국의 기밀보호 규정에 따라 공개할 수 없는 정보가 있는 경우

- DME 데이터베이스의 결측값 처리

- NULL은 자동으로 결측값으로 인식하며, 데이터 테이블에서는 '.'으로 나타남
- 식별변수(PERSID, BOOKID, SCORERID 등)에는 결측값이 없어야 함

- 결측값 정의

- 기본 설정 : 모든 데이터 레코드의 blank(NULL)는 결측값으로 정의
- 배경설문조사의 숫자형 변수의 결측값은 응답자에게서 수집한 정보와 시스템에서 생성한 정보로 구별(모름, 거절, 시스템 무응답)
- 문자형 변수(개방형 질문)에 대한 결측값은 가능한 응답의 답변내용을 그대로 유지하는 방향으로 정의

○ 기본 설정(International options)

- 인터내셔널 코드북의 참가국의 선택사항과 상관없이 프로젝트의 모든 구성요소 포함

- 인터내셔널 코드북과 데이터베이스는 구성요소의 모든 변수를 포함하지만, 결측값(NULL)으로 표시
- NDM은 해당하지 않는 변수를 삭제하거나 변경할 필요가 없으며, 국가별 변수만 추가하면 됨

라. DME software

○ 특징

- 참가국의 현장조사 이후에 데이터 관리 및 처리 지원
- 표본설계, 응답 결과, 코딩, workflow, 로그 등 관련된 데이터베이스 관리
- 인터내셔널과 내셔널 코드북 반영
- PIAAC의 다양한 소스에서 추출한 데이터의 읽기 및 통합 기능
- 채점지(scoring sheets)와 답안지(capture sheets) 내용 입력 지원
- 데이터 가져오기, 입력 시 데이터 검증
- 케이스 추가, 변수 추가 등 Merge 기능
- 입력자, 매니저, 관리자 등 권한에 따른 기능 제한

○ 시스템

- Microsoft SQL Server Compact 3.5 SP1(a SQL Server 2005/2008 spin-off for desktop use)
- Microsoft .NET Framework 3.5 SP1
- Windows XP 기반으로 설계(이후 버전 사용 가능)
- Mac, Linux 등 버추얼 환경 지원 안 함

○ 설치

- .NET Framework 3.5 설치(이미 설치되어 있을 수 있음)
- SQL Server Compact 3.5 SP1 설치(bit 런타임 버전 선택)
- DME software 설치

○ 업데이트

- 참가국은 교육 중에 배포한 첫 번째 버전과 인터내셔널 마스터 코드북으로 내셔널 코드북과 데이터 처리 업무 준비
- 참가국의 데이터 통합 테스트 과정에서 발견한 문제점을 보고하면 수정하여 SharePoint를 통해 업데이트

2. IT-C 워크숍

가. 추진 경과

- 스톡홀름 비공식 워크숍 이후 추진 경과
 - 중지(Pause) 기능 통합
 - 전체 현황 코드(Global Disposition code) 개발
 - BQ와 채점관련 문제 해결
- 통합 테스트
 - 독일에서는 스크립트를 이용한 TAO와 CMS의 통합 테스트 완료
 - 독일의 CMS : TNS Nipo(16 bit edition)
 - 중지(Pause), 재시작(ResumeCAPI)과 관련된 추가 기능에 대한 테스트 완료
- 스크립트 변경 사항('10년 2월)
 - VMware 워크스테이션과 서버 지원
 - 서버 구조의 변경 및 설정에 대한 Configuration file 추가
 - 중단된 CAPI를 재시작하는 Resume 기능 추가 : 조사를 장시간 중단한 경우나 조사 완료 후 조사원이 면접관찰설문을 응답할 때 필요

나. 결과

- 국내 CMS 개선 방안
 - 조사용 노트북에 녹음기능 추가 : 예비조사에서는 적용 불가하며, 본 조사에서 CMS에 녹음 기능 추가 여부 검토
 - 예비조사에서는 무선인터넷을 활용할 수 없어 모든 조사결과는 USB를 통해 저장하지만, 네트워크 연결이 가능한 경우 활용할 수 있도록 조사원이 직접 업로드 할 수 있는 기능 추가 필요
 - 문제가 발생하여 시스템 재설치가 필요한 경우를 대비하여 설치 파일을 각 노트북에 저장(파일 접근 권한은 관리자 권한으로 제한)

Ⅲ. 향후 추진 일정

- 시스템 준비
 - DME(Data Management Expert) 테스트 : '10. 4. ~ '10. 5.
 - 조사용 노트북 임대 및 설치 테스트 : '10. 4.
- 교육·훈련 자료 준비
 - 품질관리 지침서 및 보고서 양식 작성 : '10. 2. ~ '10. 3.
 - 조사원 교육 PPT, Script 작성 : '10. 2. ~ '10. 4.
 - 조사용품 제작 : '10. 2. ~ '10. 4.
- 2010~2013년 주요 업무 일정
 - 예비조사 담당자 교육 : '10. 4.
 - 예비조사(Field Test) : '10. 5. ~ 6.
 - 예비조사 결과 처리 및 제출 : '10. 7. ~ 8.
 - 예비조사 결과 분석(컨소시엄, 참가국) : '10. 9. ~ 11.
 - 본 조사를 위한 6차 NPM 회의 : '10. 12.
 - 본 조사를 위한 표본설계 워크숍 : '10. 12.
 - 본 조사 National VM 개발 : '11. 1. ~ 2.
 - 본 조사 평가 도구 수정 완료 : '11. 3. ~ 5.
 - 본 조사 international 교육/훈련 : '11. 6.
 - 본 조사(Main Study) : '11. 8. ~ '12. 3.
 - 본 조사 결과 분석 및 발표 : '12. 4. ~ '13. 9.