Norman Wright

The need for high volume, low cost data entry is leading increasingly to the use of optical mark sense equipment in a wide variety of applications. Frequently the greatest part of the effort of the beginning user goes into assessing the capabilities of the various available OMR media. Yet in most applications, this will represent only the opening pages of OMR implementation. This discussion will concentrate on such factors as the impact of forms design on the accuracy and completeness of OMR data. Consideration will also be given to philosophies of error correction as well as the implications of OMR data entry on overall systems design. The main thrust of the discussion is on the non-mechanical aspects of humanely and accurately "marking sense".

"The Moving Finger writes: and having writ, moves on: nor all thy piety nor wit shall lure it back to cancel half a line, nor all thy tears wash out a word of it." (From The Rubaiyat of Omar Khayyam, Edward Fitzgerald, translator.)

When the poet Omar Khayyam penned these words in 12th Century Persia, he did not rule out the possibility that, while we could not change a line of what is writ, we could at least subject it to intensive study and re-interpretation. And while Omar could not have foreseen the modern corollary, he almost certainly would have understood it: to capture every line--preferably in machine-readable form. This modern need for high volume data entry leads us, not completely by punsmanship, to our modern OMR--the optical mark reader.

Historically the optical mark reader is closely associated with educational applications. One of its first uses was in scoring the large volumes of standardized educational tests so familiar to post-war generations of students. At first mechanical, and later optical mark sensing equipment freed countless teachers and clerks from the tedious, time consuming, and error-prone task of hand scoring the thousands of student examinations that went along with mass assessments of student achievement. The almost immediate acceptance of the mark sense response sheet may be attributed to its reliance upon the simple and familiar tools of paper and

pencil, and to its close analogy to its non-machine-scorable counterpart. The OMR has continued to dominate this field. Tests, surveys, and questionnaires of all kinds today remain the most widespread and familiar of all OMR applications.

Because of its use in education, it is not surprising that educational administrators were one of the first groups to see the value of the technology in gathering demographic data from students. In schools and colleges around the country, OMR data entry was soon being used not only to assess students, but to register them, assign them to classes, and report their final grades as well. This natural succession of events was perhaps accentuated by the fact that OMR did best what educators needed to do--to capture large amounts of data in a very short time at minimal cost. These non-testing uses of the OMR helped form a broader base upon which many more recent applications have been built.

Increasingly optical mark sense technology is finding its way into diverse applications in business, industry, and government. It has been used to capture census, personnel, and survey information closely analogous to its uses in education. It has also been used for many traditional applications in business--entering routine statistical and sampling data, quality control, inventory, and time accounting. Recently the OMR has been used to replace hand-counted paper ballots in a voting system where the legal requirements of ballot layout prohibited conventional voting machines.

OMR media today are available in a wide variety of styles and sizes. Documents range from punched card size, to standard 8½" by 11" sheets, to documents which fold out into a long continuous form. Many of the readers for card-sized documents accommodate punched as well as marked information.

What all of the available media have in common is requiring the respondant to record his data by gridding--filling in pre-printed boxes, ovals, or circles with pencil marks. The gridded marks are read by the photo sensors on the mark reader and translated by hardware and software into dicital data. The sample sheet

(fig. 1), used by the American College Testing Program to register students for its examinations, shows some of the variety of information that can be captured in this manner. This is in addition, of course, to the more usual multiple choice type of responses.

The example is notable for the relatively high amount of written and numerical information which the applicant is requested to grid. Unfortunately, no studies have been done, as they have been for CRT's, testing the psychological attitude and tolerance of respondants to this form of data entry. Unquestionably the gridding of the information takes longer than printing it--about 3 minutes for gridding the sample document, versus about 1½ minutes for simply entering the data in printed form. In some sense at least, the respondant is required to function as a data transcriber as well as to provide the required information. Undeniably this form of data entry depends upon the time, effort, and care of the respondant.

Largely for this reason it was once felt that filling out OMR documents should take place only in carefully supervised environments. While this is still frequently practiced in schools and colleges, many OMR users have recently had good results with documents filled out by the general public in uncontrolled situations. The sample document in figure 2 is designed for home use. Users of data sheets and surveys find that they can usually rely upon the respondant's good will and his desire for personal data to be correctly entered. There is also some evidence to suggest that the present population consists increasingly of "trained" data transcribers, since OMR is becoming widespread.

In applications requiring use by the untrained respondant or the general public, the layout and design of the printed form is crucial to the success of any OMR project. Essential to successful layout is some ordered and clear method of presenting data for response. With sheet sized documents, the user can frequently include explanations and examples on the OMR document itself. If the information is not to be filled out in a controlled environment, it is almost always desirable to include accompanying instructions clearly specifying the desired content of each

field. The material should also give the traditional "Grid only one response" instructions for marking OMR format; however, too frequently this is done to the detriment of elaboration on field content.

Almost all OMR projects require that related documents be grouped into some form or other, whether by classes and schools in educational applications, or by production units in a business environment. The older method for entering such control information was to depend on the respondant to enter it correctly on the response document itself. For example, students would be asked to grid a school number and the first 5 letters of their teacher's name. Increasingly, however, OMR users prefer to enter such control information on separate sheets or documents, usually ones which are identifiable to the scanning software or hardware itself. Because of its pivotal importance control data is thus separated from respondant data. Frequently it may be entered by a more highly trained or carefully controlled person--a teacher or a clerk at the processing center.

At the U.S. Civil Service Commission's OMR processing center, for example, written exams are shipped to the center from 65 area offices around the country. Personnel at the center fill out a scannable document known as an Area Office Header to identify each group of answer sheets received. The header sheet is color coded and corner cut, borrowing techniques from card processing. This header itself becomes a record on tape along with job applicant documents. It identifies all subsequent records on tape as being input from a particular area office, as well as giving processing and control information about the documents.

In the simplest of all imaginable mark sense applications incorrectly gridded data is of two sorts: either an applicant grids more than one response on a set of grid positions requiring only one--known as a set of grid positions requiring only one--known as a double grid; or the respondant fails to grid any response—an omit. In some applications, the software, either at the point of scanning or in subsequent processing, can be programmed to predictably circumvent or default when

encountering such data. More usually the data suspected to be in error will be examined to see if some determination can be made of the intent of the respondant. Double grids and omits give greatest problem in demographic data; fortunately it is also here that they are most likely to be correctable. Almost universally, OMR documents requiring gridded data such as name and address request that the respondant both print and grid his information. In such cases most errors can be resolved by resorting to the written information.

Some users of OMR depend on extensive pre-screening of the mark sense documents in an effort to detect gridding errors before they are input to the machine. Such an examination can never be totally effective, of course. More importantly, it tends to use a great deal of human time and energy in perusing documents which will need no attention. For this reason, it is usually much more cost effective to concentrate on errors after they have been detected by computer.

In most applications, then, the raw data from the scanner will pass through some kind of edit program designed to detect errors peculiar to OMR entry. Usually other edits and checks on data validity, similar to those for other methods of data entry, will also be performed. The resolution and correction of the detected errors will usually represent by far the greatest amount of time and effort spent in any system involving OMR entry.

The correction process can be considerably more complex using OMR technologies where only one side of a document can be read at a time, or where a single respondant's data spans several documents or cards. In such applications where the order and relatedness of documents can usually not be depended on, a match is done to bring together the different pages or documents which form data on a single individual. Many schemes for pre-gridding or pre-punching match information have been devised. Unless some provision is made for strict control on the match fields, resolution of mis-matching pages can be a monumental task.

The actual method by which corrections are applied to files of OMR data presents an interesting paradox. Since the data is already in machine readable form on the

mark sense document, and since it can usually be corrected by simply erasing or re-gridding the mis-gridded information, it would seem natural to apply the corrections simply by re-scanning the corrected document. While some OMR users do opt for this method, most high volume users do not. They object that it disrupts the production flow of batches and complicates the process of undating data with corrected information. They argue that, while OMR is good at the task of rapid mass data entry, keyed entry is more suitable to correcting the data once it is at a data center.

In an average, well designed OMR application, the documents containing errors or suspected errors will represent somewhere between 5 and 20 percent of the total number of documents entered. The exact figures will, of course, depend on many factors including the amount of data on each document, the age and experience of the respondants, and the complexity of the edit criteria themselves. Typically, the number of actual corrections to be applied to the data will be sonderably less than the number of documents examined for suspected error. Most systems utilizing OMR entry today provide for field corrections on erroneous data rather than requiring total record replacement. This design factor brings about important economies in the data entry phases of correction. The percentage of fields in error, of course, is drastically less than the percentage of documents in error.

Recent advances in CRT technology, and the cost reductions which have accompanied them, open broad new avenues for error detection and correction in OMR data. A few of the newest systems have been designed utilizing some form of CRT correction entry. One can envision the entire process of error detection and correction taking place at the CRT. The stack of OMR documents would provide the only source material needed for the correction process. In such a set up, a single analyst would be responsible for both the resolution of the error and the actual correction of the data on file.

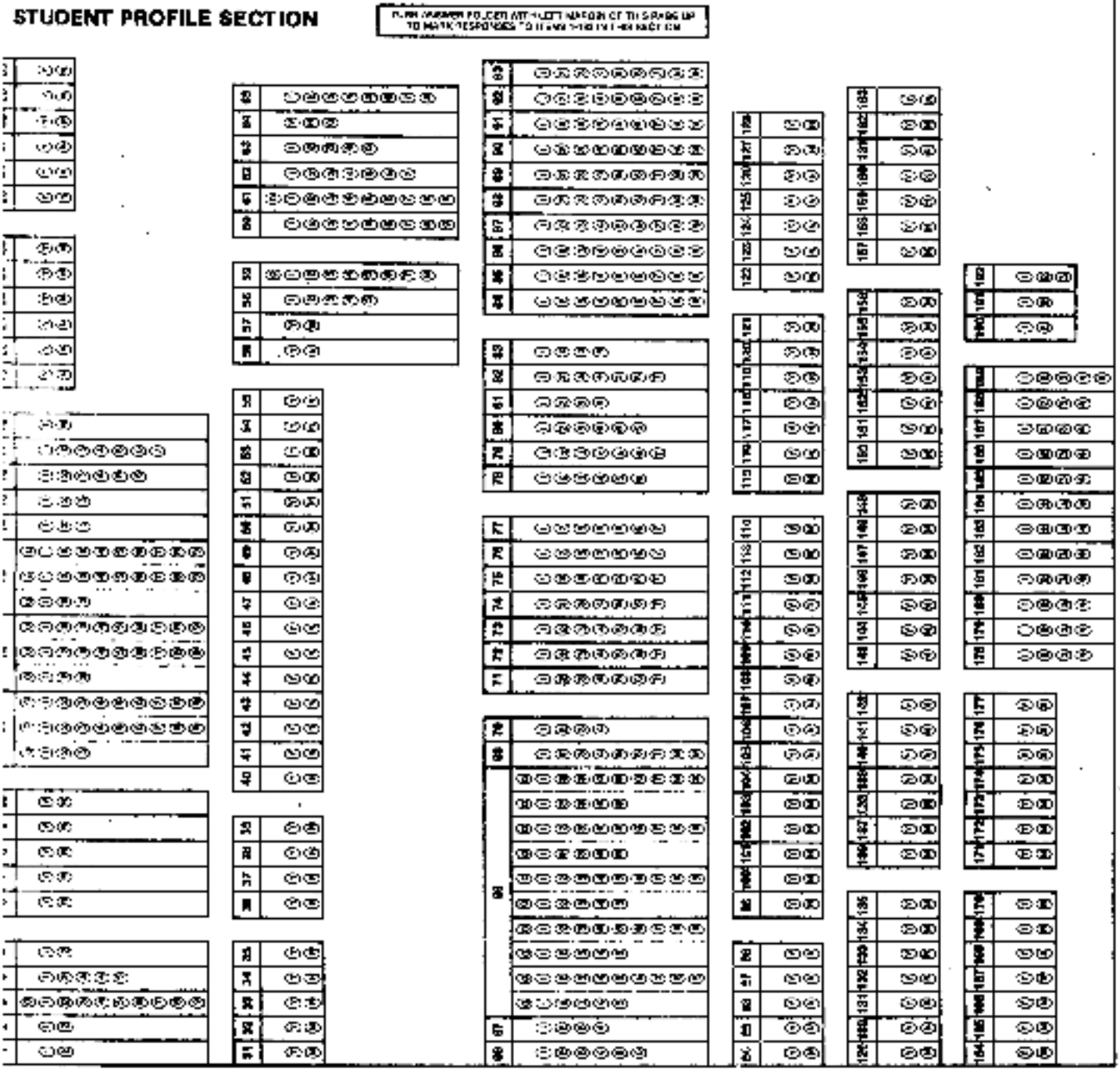
Regardless of their differences in application and technique OMR users point to many of the same advantages. Data is captured in machine readable form directly

from the respondant, thus eliminating the time consuming task of keyed entry. The CMR document is highly portable and requires no special equipment to fill out. While data gathered in this manner will almost always contain errors, a well-designed system can detect and recover from such errors effectively. OMR's are highly efficient at entering large masses of data at low cost. Thus in many applications involving information from and about people, optical mark readers will continue to represent one of the best means available for data aquisition.

INTEREST INVENTORY

1	13	25	37	49	61	73	85
2	14	26	38	50	62	74	86
3	15	27	39	51	63	75	87
4	16	28	40	52	64	76	88
5	17	29	41	53	65	77	89
6	18	30	42	54	66	78	90
7	19	31	43	55	67	79	
8	20	32	44	56	68	80	
9	21	33	45	57	69	81	
10	22	34	46	58	70	82	
11	23	35	47	59	71	83	
12	24	36	48	60	72	84	

STUDENT PROFILE SECTION



MMC FORM NO. 6301

U.S. Civil Service Commission

MID-LEVEL DATA SHEET

FORM APPROVED
OMB NO. 50-R0437
FORM **B**

GENERAL INSTRUCTIONS: The questions on this form request information about your background, interests, and availability. Before you make any entries, make sure you have read all instructions on this form, accompanying Form 1156A, and any other instructional material provided by the area office with this form. Recording information that does not accurately reflect your availability, employment preferences, and qualifications may cause you to lose consideration for some positions.

This form can only be processed if you -- (1.) Use a number 2 (or softer) lead pencil. (2.) Completely blacken the oval corresponding to your response choice. (3.) Completely erase any mistakes. (4.) Make no stray marks.

NOTE: BE SURE TO WRITE IN YOUR RESPONSE AS WELL AS BLACKENING THE OVALS IN THE BLOCKS WHERE BOTH ARE REQUIRED.

1 NAME: _____ (Please Print)

DATE: _____ (Please Print)

2 SOCIAL SECURITY NUMBER
Record your SSN in the spaces provided. Blacken the appropriate ovals below.

3 GENDER
Blacken the appropriate oval.
○ Male ○ Female

4 GRADE LEVEL(S)
Refer to the enclosed qualifications guidelines. Blacken the oval(s) beside the grade level(s) for which you are available and believe you are qualified.
○ GS-9 ○ GS-11 ○ GS-12

5 TRAVEL
Are you willing to accept a position that requires travel away from the city of your normal work location at least 25 percent of the time?
○ Yes ○ No

6 TEMPORARY EMPLOYMENT
Are you willing to accept temporary employment?
○ Yes ○ No

7 VETERAN PREFERENCE
Refer to the Form 1056B instructions. Blacken the oval that corresponds to your preference claim.
○ Disability 10-pt (CP) ○ 5-pt (TP)
○ Other 10-pt (XP) ○ None (NV)

8 PART-TIME EMPLOYMENT
Are you willing to accept part-time employment (fewer than 40 hrs. per week)?
○ Yes ○ No

9 EDUCATION LEVEL
Blacken the oval beside the highest education level you have attained, or will attain in the next 9 months. Only one oval may be blackened.
○ Ph.D
○ Masters
○ LL.B. or J.D. Degree
○ BA Degree
○ Associate Degree
○ Some College
○ High School or Equivalency
○ Less than High School

10 LANGUAGES
Blacken the oval beside each language (other than English) which you read and speak fluently.
○ French ○ Spanish
○ Russian ○ Chinese
○ German ○ Other

11 FEDERAL EMPLOYEE
Are you now a Federal employee?
○ Yes ○ No

12 STATE/LOCAL AVAILABILITY
Are you interested in being considered for State or local Government employment?
○ Yes ○ No

Page 2

193

13 STATE OF LEGAL or VOTING RESIDENCE

Blacken the oval beside the State of your legal or voting residence. Only one oval may be blackened.

- ○ Alabama
- ○ Alaska
- ○ Arizona
- ○ Arkansas
- ○ California
- ○ Colorado
- ○ Connecticut
- ○ Delaware
- ○ District of Columbia
- ○ Florida
- ○ Georgia
- ○ Guam
- ○ Hawaii
- ○ Idaho
- ○ Illinois
- ○ Indiana
- ○ Iowa
- ○ Kansas
- ○ Kentucky
- ○ Louisiana
- ○ Maine
- ○ Maryland
- ○ Massachusetts
- ○ Michigan
- ○ Minnesota
- ○ Mississippi
- ○ Missouri
- ○ Montana
- ○ Nebraska
- ○ Nevada
- ○ New Hampshire
- ○ New Jersey
- ○ New Mexico
- ○ New York
- ○ North Carolina
- ○ North Dakota
- ○ Ohio
- ○ Oklahoma
- ○ Oregon
- ○ Pennsylvania
- ○ Puerto Rico
- ○ Rhode Island
- ○ South Carolina
- ○ South Dakota
- ○ Tennessee
- ○ Texas
- ○ Utah
- ○ Vermont
- ○ Virginia
- ○ Virgin Islands
- ○ Washington
- ○ West Virginia
- ○ Wisconsin
- ○ Wyoming

14 PREFERENCE

The purpose of this item is to ask you to identify the state or area where you most prefer to work. Please refer to the local area office's mid-level information sheet. State of preference will only be used for well qualified candidates for that location. Place the code you select in the spaces provided. Blacken the appropriate ovals.

15 SPECIAL SKILLS and KNOWLEDGES

Listed below are nine special skills and knowledges related to positions at the level for which you are applying. You may select up to four of those skills and knowledges for which you have applicable experience and/or training. *Do not select an item if you would not be interested in a job which requires that special skill or knowledge.* Blacken the oval beside each choice.

- ○ **Program planning or development** – systematically applying broad program goals to develop new or improved programs.
- ○ **Program evaluation** – evaluation of ongoing programs to improve their effectiveness. Identifying shortcomings and recommending alternative approaches.
- ○ **Program management** – directing and controlling programs. Deciding content, objectives and priorities, and allocating organizational resources.
- ○ **Supervising** – planning and directing the work of other employees.
- ○ **Preparation of Written Reports, Program Proposals.**
- ○ **Adapting Operations and Procedures to ADP Processes.**
- ○ **Teaching/Training** – serving as an instructor at training sessions. Involves preparation of course material.
- ○ **Grants Review** – reviewing grant applications and determining grant awards.
- ○ **Knowledge of Engineering and Scientific Fundamentals.**

16 OCCUPATIONAL CODES

Please refer to the local area office's mid-level information sheet. Record the codes you choose in the spaces provided. Blacken the appropriate ovals below.

17 COLLEGE MAJOR

If you have a college degree (or expect to receive one within the next 9 months), blacken the oval corresponding to your major field of study. If you are qualified on the basis of education alone, you must have at least 2 years of graduate study or a Master's Degree. Only one oval may be blackened.

- ○ Accounting
- ○ Anthropology
- ○ Agriculture, Fish and Wildlife Management, Park Management, Forestry
- ○ Archeology
- ○ Biological Sciences
- ○ Business Administration
- ○ Communications
- ○ Computer Science
- ○ Criminology
- ○ Economics
- ○ Education
- ○ English, Journalism
- ○ Engineering (except Industrial)
- ○ Fine or Applied Arts
- ○ Foreign Languages
- ○ Geography
- ○ History
- ○ Industrial Engineering
- ○ Law
- ○ Library Science
- ○ Mathematics or Statistics
- ○ Personnel or Industrial Relations
- ○ Pharmacy
- ○ Physical Science
- ○ Political Science, Public Administration
- ○ Psychology
- ○ Public Health
- ○ Social Work or Sociology
- ○ Social Sciences (other than those listed)
- ○ Transportation
- ○ Urban Planning
- ○ Other

18 GEOGRAPHIC AVAILABILITY

Refer to Form 1056B instructions. Blacken the ovals beside your choices.

○ 01	○ 11	○ 21	○ 31	○ 41
○ 02	○ 12	○ 22	○ 32	○ 42
○ 03	○ 13	○ 23	○ 33	○ 43
○ 04	○ 14	○ 24	○ 34	○ 44
○ 05	○ 15	○ 25	○ 35	○ 45
○ 06	○ 16	○ 26	○ 36	○ 46
○ 07	○ 17	○ 27	○ 37	○ 47
○ 08	○ 18	○ 28	○ 38	○ 49
○ 09	○ 19	○ 29	○ 39	
○ 10	○ 20	○ 30	○ 40	

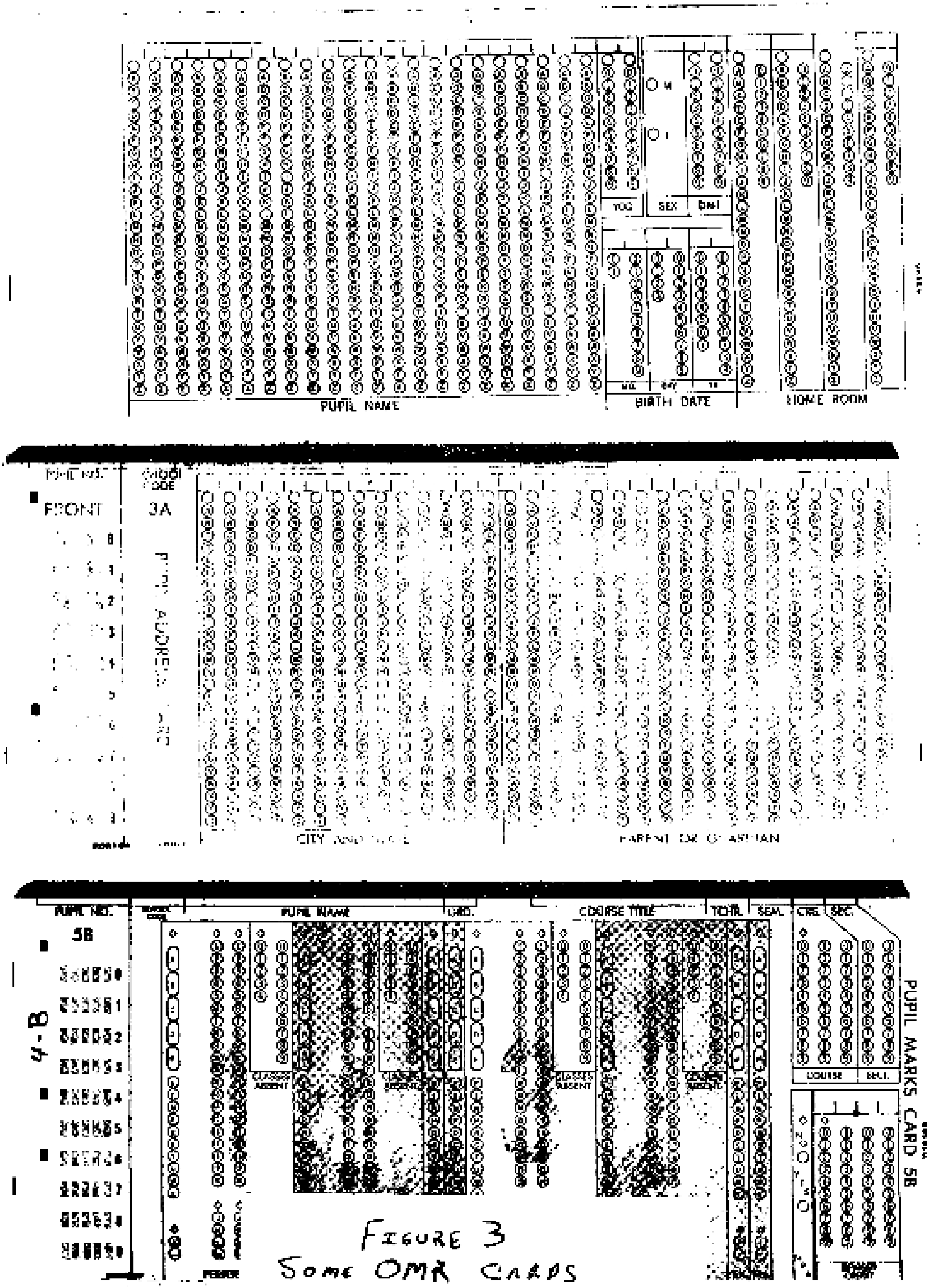


Figure 3
Some OMR Cards