

18th Population Census Conference  
26-29 August 1998, Honolulu, Hawaii, USA

**Cost-effective Approaches to  
the 2000 Population and Housing Census  
in Korea**

by

**Sang-Sik Kim**  
Director, Population Census Division

and

**Hyung-Seog Kim**  
Demographer, Population Census Division

National Statistical Office  
Republic of Korea

## **. Introduction**

1. This paper aims at sharing the experiences acquired in the process of conducting pre-tests to examine cost-effective approaches to the population and housing census in Korea, referring to South Korea if not stated. The findings of pre-tests taken in 1997 and 1998 serve the major materials for this paper.
2. Korea has conducted population and housing censuses on a quinquennial basis since 1925. The censuses have been of fundamental statistics not only for policy-making, socioeconomic development planning but also for administrative, research and business purposes. In process of socioeconomic developments in Korea, the censuses have well acknowledged to be primarily useful source to plan, implement and evaluate national and local development programs.
3. It is, however, not likely to hold true any more that the census should be conducted only because it is important and has been carried out historically as well as universally in the world. Due to huge amount of census-taking costs together with recent financial difficulties in Korea, it is predicted to be very difficult to ensure due priority and adequate resources required for the 2000 Population and Housing Census. It would be never easy task for census officials to persuade people and budget authority of the justifiability of the census-taking.
4. Considering the increasing trends of wage rates and size of population, the amount of costs for the 2000 Census is expected to reach around 100 billion *won* which is almost twice as much as that for the 1995 Population and Housing Census. Therefore, it is a priority matter that we have to think over various ways to reduce increasing costs for the 2000 Census in Korea.
5. Census operations have been taken through long procedures; preparatory work, enumeration, data capture, data processing and data dissemination. All of these steps should be scrutinized whether they are well planned, prepared and operated in an effective and efficient manner. Minor ineffectiveness and inefficiency in the operations at each step may lead to waste of available resources. In this regard, the methodology and techniques employed at each stage of the census are of great importance.
6. On the other hand, the user's demand for census data has been diversified not only in contents but also in media. They want to access the census data in cheaper and more convenient manner. It should be taken into account in the context of cost-benefit analysis to increase the value of census data through meeting various demands.
7. Strategies for the cost-effective approaches to the 2000 Population and Housing Census in Korea are focused, firstly, on the adoption of a self- enumeration method in principle instead of traditional face-to-face method employed so far and, secondly, on the introduction of an OCR technique in data capturing.
8. The feasibility of these approaches in terms of cost, management and operation are currently under study. Since the fundamental changes in methodology are being considered, several detailed preliminary studies and field tests are required. Pre-tests were carried out in October 1997 and June 1998.

## **. Cost-effective Approaches Under Considerations**

### ***Improving Method of Enumeration***

The method of enumeration can be generally dichotomized into a self-enumeration method and a canvasser method. The latter refers to the method that enumerators visit every household for a face-to-face interview while the former refers to the method that household members fill out questionnaires administered to them in advance.

10. With reference to the self-enumeration method, several different approaches are practicable, that is, mail-out/mail-back, delivery/collect, mail-out/collect, delivery/mail-back methods. During the period of 1985-1994, 163 countries have reportedly conducted the population and housing censuses, and the canvasser method was employed by 124 countries, the mail-out/mail-back method by 13 countries, the mail-out/collect-back method by 3 countries and the delivery/collect method by 34 countries respectively (Yu, 1995).

11. We, so far, have long employed the traditional canvasser method in principle but mixed with the self-enumeration method in practice; questionnaires may be distributed to respondents if they want to fill out the forms by themselves and, also, questionnaires may be indirectly distributed via their neighbors if enumerators fail to meet respondents face-to-face. Since the self-enumeration method used to be partially practiced in fields, the census questionnaires have been designed in consideration not only of face-to-face interview but also of self-enumeration.

12. The pre-tests were carried out twice in October 1997 and June 1998 to study the quality of data, the cost and the operational feasibility of introduction of self-enumeration method. The immediate objectives of the tests are to measure the percentage of households to which the delivery/collect method is applicable as well as the accuracy of questionnaires replied by the same method. Samples were drawn on a purposive basis, considering the level of urbanization and of housing units. The sample size of the 1997 pre-test was 650 households and that of the 1998 pre-test was 7,800 households out of the whole 12,958 thousands households.

13. It was revealed in the 1997 pre-test that the delivery/collect enumeration method is unlikely to be applicable unanimously to every household due to the inability in response among the disadvantageous groups like the elderly, the blind and the illiterate and so on. The application of delivery/collect method is suggested to be selective in terms of the level of urbanization and of the type of housing units. The proportion of households which were able to be enumerated by the delivery/collect method in the 1997 pre-test was 91.3 percent whereas the proportion by the canvasser method was 8.7 percent. Breaking down into the level of urbanization, however, rural areas were apparently low with 66.1 percent in the proportion of households enumerated by the delivery/collect method. In general, one out of three households had to be interviewed by enumerators due to the elderly in rural areas. In terms of type of housing units, apartment marked 96.4%, a slightly higher percentage in the proportion of self-enumerated households than other types of housing units(refer to Table 1).

14. As far as the accuracy of replies is concerned, the delivery/collect method appeared to be more applicable to the short-form of questionnaire than to the long-form questionnaire. The eye-checking of questionnaires in the 1997 pre-test showed that, in short-forms, only 30.2 percent of questionnaires

turned out to be free from post-corrections while 19.5 percent of questionnaires were required to be corrected but without referring to the respondents and 50.3 percent needed to be asked additional inquiries by either telephoning or revisiting; whereas, in long-forms, only 16.5 percent of questionnaires were correctly answered, free from post-corrections, while 11.8 percent to be corrected without referring to the respondents. The rest 71.7 percent needed to be referred to the respondents for the supplementary information. In terms of type of housing units, the highest proportion of questionnaires free from additional corrections is appeared in the apartment households with short-forms, reflecting the homogeneity of housing conditions among them. On the other hand, the households in detached dwelling marked 86.8 percent which is higher proportion of questionnaires to be corrected with additional field inquiries than other two types of households(refer to Table 2).

Table 1. Number of Households by Enumeration Method: In the 1997 Pre-test for the 2000 Population and Housing Census, Korea

Note : Figures in parentheses show the proportion of households by enumeration methods

	No. Of House-holds	Response (in households)					Non-response (in households)		
		Total	By enumeration method				Total	due to	
			Delivery /collect	Cavasser		Absent		Refusal	
				Sub-Total	due to				
					Old age	Others			
Total	650	586(100.0)	535(91.3)	51 (8.7)	35	16	64	25	39
a. Levels of urbanization									
) Metropolitan cities	349	311(100.0)	299(96.1)	12 (3.9)	4	8	38	16	22
) Small cities	239	219(100.0)	199(90.9)	20 (9.1)	12	8	20	8	12
) Rural areas	62	56(100.0)	37(66.1)	19(33.9)	19	-	6	1	5
b. Housing units									
) Detached dwellings	281	259(100.0)	229(88.4)	30(11.6)	25	5	22	6	16
) Apartments	187	165(100.0)	159(96.4)	6 (3.6)	6	-	22	15	7
) Row houses	182	162(100.0)	147(90.7)	15 (9.3)	4	11	20	4	16

15. On the other hand, the number of households to which inquiries were made for correction were 103 households by telephone and 20 households by revisit out of 535 households enumerated by the delivery/collect method; wherein, 3.2 items and 5.0 items were asked on an average by telephone and by revisit respectively.

16. The proportions of households whose replies to certain question were correct out of total households enumerated in the delivery/collect method were 98.7% for name, 95.7% for sex and also for relationship to head of household, 93.5% for type of living quarter whereas 50.4% for whether of commuting, 56.5% for place of birth and also for place of residence prior to 5 years, 62.2% for type of household, 63.6% for occupation of worker and 65.6% for industry of worker.

Table 2. The Accuracy of Questionnaires Enumerated by Delivery/Collect Method:  
In the 1997 Pre-test for the 2000 Population and Housing Census, Korea

	Total	Detached dwelling	Apartment	Row house
<b>A. Short-forms <sup>1</sup></b>				
a. Total no. of households	298(100.0)	131(100.0)	81(100.0)	86(100.0)
b. No. of households whose questionnaire is				
) free from correction	90 (30.2)	36 (27.5)	30 (37.0)	24 (27.9)
) to be corrected without additional field inquiries	58 (19.5)	9 (6.9)	26 (32.1)	23 (26.8)
) to be corrected with additional field inquiries	150 (50.3)	86 (65.6)	25 (30.9)	39 (45.3)
<b>B. Long-forms <sup>1</sup></b>				
a. Total no. of households	237(100.0)	98(100.0)	78(100.0)	61(100.0)
b. No. of households whose questionnaire is				
) free from correction	39 (16.5)	6 (6.1)	18 (23.1)	15 (24.6)
) to be corrected without additional field inquiries	28 (11.8)	7 (7.1)	15 (19.2)	6 (9.8)
) to be corrected with additional field inquiries	170 (71.7)	85 (86.8)	45 (57.7)	40 (34.4)

Notes : 1. The short-form has 17 items while the long-form has 28 items.

2. Figures in parentheses show the percentages to total no. of household in each panel.

17. The average frequency of visits per household was 1.3 times for the delivery of

questionnaires and 2.1 times for their collection. The average time spent per a single household for questionnaire delivery was 3.9 minutes while that for questionnaire collection was 4.6 minutes. Including 5.8 minutes for telephoning and revisiting to correct errors committed by householders, total average enumeration time per a household amounted to 16.3 minutes.

18. Assuming 6 hours as working time a day, one enumerator is presumed to be able to work on around 20 households per a day by the delivery/collect method. Accordingly, within 9 days given for enumeration in the census, 3 EDs(Enumeration Districts), wherein an ED comprises about 60 households, are assumed to be adequate workload per an enumerator for the short-form questionnaire instead of 2 EDs allocated to one enumerator in the past. Similarly, the workload of 2 EDs per an enumerator is feasible for the long-form questionnaire instead of 1 ED before.

19. It furnishes a key to the successful introduction of delivery/collect method to reduce considerably errors both on coverage and on contents. In this regard, the delivery/collect method is suggested to be selectively applied to the households in apartments, whose households' number is estimated to reach around 6 millions in the year 2000, because of the homogeneity of their living conditions; since one apartment living quarter has by and large one household unlike other types of living quarters, there are less probabilities to omit counts and their similarity in housing conditions is useful for quality control in such questionnaire items as total area of floor space, number of rooms, year of construction and number of housing facilities and so on. Apartment households are also very easy and convenient in distributing and collecting questionnaires since they are located within the same building and have security guards in many cases. It would also help to shoot at such targets to design questionnaires in a respondent-friendly manner.

20. In the 1995 Census in Korea, the number of enumerators employed was around 125 thousands, accounting for more than 85 percent of total costs of census-taking. By increasing the number of EDs per an enumerator in the 2000 Census, around 50 thousands enumerators are estimated to be curtailed, resulting in saving 18 billion *won*. Even though the application of delivery/collect method should be, considering strictly the accuracy of replies, limited to only apartments, about 20 thousands enumerators and 6 billion *won* are also expected to be saved.

### ***Introduction of Image Data Capture***

21. Before the 1990 Census, Korea employed keyboard entry method on a outsourcing basis in data capturing. The OMR technique was introduced first to the census in 1990 to reduce the entry errors and to shorten the data-processing periods. The trend of increasing labor shortage in key-entry industry was also taken into consideration. Generally speaking, the introduction of OMR technique to the population and housing census was quite successful in the Korean context. The production of paper was localized for the 1995 Census. Owing to the OMR system, the period of data-processing were largely curtailed from two years to one year, resulting in encouraging fast data dissemination.

22. As OMR questionnaires were, however, so sensitive to contamination and damage, the two types of questionnaires, that is, ordinary forms and OMR forms, had to be simultaneously used in 1990 and 1995. And, converting manually completed ordinary

questionnaires to OMR forms became a major source of errors.

23. The out-of-dateness of OMR machines purchased in 1989 and 1990 raises the critical questions regarding the feasibility of the OMR data capture in the 2000 Census. The repairs of OMR machines expect to be increased sharply not only in frequency but also in degree; they were downed 65 times and thus repaired 18 times in the process of reading questionnaires in the 1995 Census. In particular, the production stoppage of several components, for instances, MARK HAED ASSY and OMR AMP PKG and so on, threatens deeply the successful data processing for the 2000 Census. The supply of such components is not likely to be possible at this moment. In addition, the inappropriateness of OMR format questionnaire to the delivery/collect enumeration method is brought into another question since the handling of such kind of questionnaire in fields is not easy task at all.

24. The NSO of Korea has started to take it into consideration to adopt a so-called OCR or image data-capturing in the hope that this technology would allow census-takers to print only one type of questionnaire instead of two types in the OMR system, resulting in eliminating the errors committed in transcription as well as in reducing costs in printing questionnaires. Since the questionnaire for OCR is not necessarily required to be printed in special quality as does for OMR. It has an considerable tolerance to variations in paper quality.

25. The 1998 pre-test was carried out in June this year, by using OCR format questionnaires. The OCR system is comprised of five stages of processing, that is, scan station, filtering station, recognition station, correction station and post-processing station. Major features of each station are: ) questionnaires are scanned in scan station; ) broken strokes are connected and noisy dots are removed in filtering station; ) rejection marks are designed to appear in red for manual confirmation when there is doubt on recognition in recognition station; ) in the correction station, numerals and marks are confirmed by eye-check and mis-recognition are to be corrected; and, ) logical checks and editing including imputation based on other entries on the questionnaire are made in post-processing station(Figure 1).

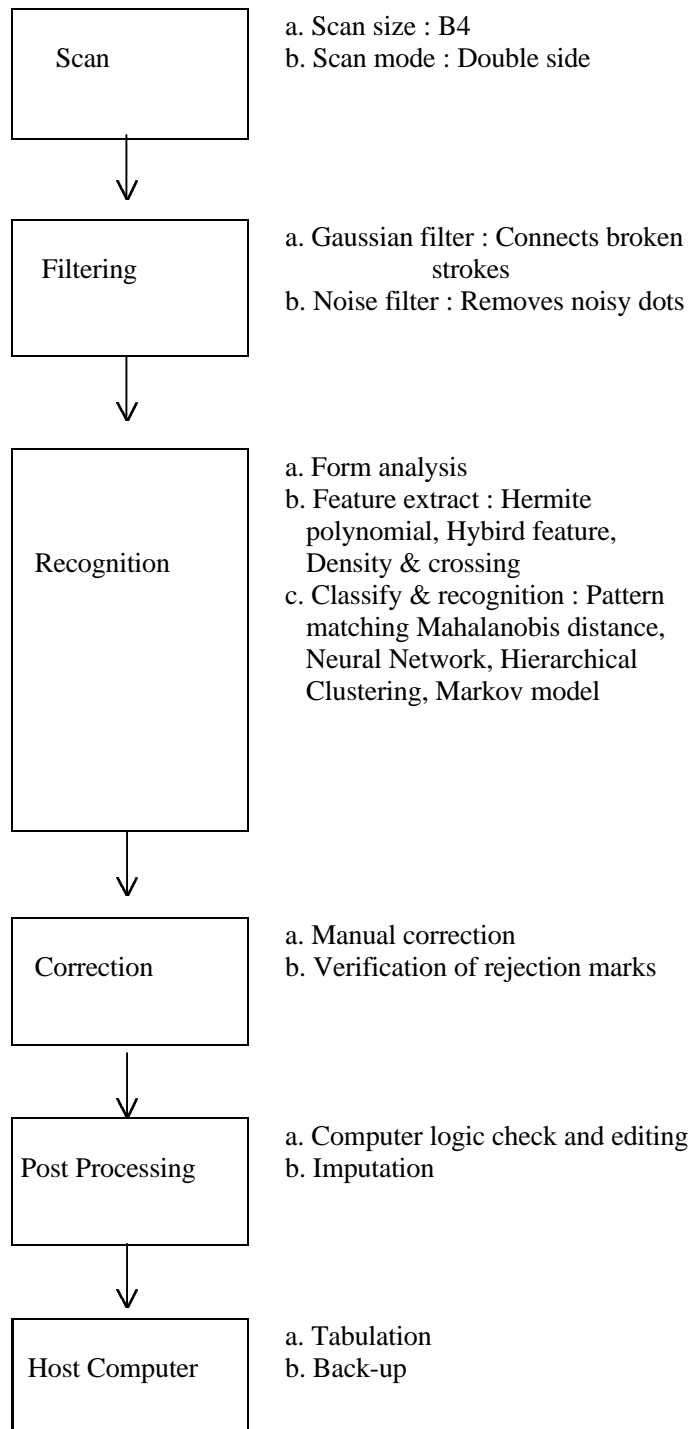


Figure 1. Flow Design of OCR Data Capture at the 1998 Pre-test for the 2000 Population and Housing Census in Korea

Table 3. Recognition Rates in OCR Format Questionnaires: In the 1998 Pre-test for the 2000 Population and Housing Census, Korea

	Number of recognition					
	Total	CWO	CW	IW	IWO	NWO
Total	804 123 (100.0)	756 539 (94.1)	31 989 (4.0)	10 720 (1.3)	4 280 (0.5)	595 (0.1)
a. Types of recognition						
) Numerals	463 710 (100.0)	419 168 (90.4)	31 619 (6.8)	10 128 (2.2)	2 387 (0.5)	408 (0.1)
) Marks	340 413 (100.0)	337 371 (99.1)	370 (0.1)	592 (0.2)	1 893 (0.6)	187 (0.1)
b. Questionnaires						
) Long-forms	357 104 (100.0)	336 012 (94.1)	14 126 (4.0)	4 317 (1.2)	2 421 (0.7)	228 (0.1)
) Short-forms	447 019 (100.0)	420 527 (94.1)	17 863 (4.0)	6 403 (1.4)	1 859 (0.4)	367 (0.1)
c. Levels of urbanization						
) Metropolitan cities	185 423 (100.0)	176 817 (95.4)	6 050 (3.3)	1 844 (1.0)	656 (0.4)	56 (0.0)
) Small cities	424 416 (100.0)	398 160 (93.8)	17 656 (4.2)	5 886 (1.4)	2 362 (0.6)	352 (0.1)
) Rural areas	194 284 (100.0)	181 562 (93.5)	8 283 (4.3)	2 990 (1.5)	1 262 (0.6)	187 (0.1)
d. Housing units						
) Detached dwellings	264 324 (100.0)	246 802 (93.4)	11 648 (4.4)	4 028 (1.5)	1 629 (0.6)	217 (0.1)
) Apartments	292 153 (100.0)	278 281 (95.3)	9 436 (3.2)	3 006 (1.0)	1 280 (0.4)	150 (0.1)
) Row houses	247 646 (100.0)	231 456 (93.5)	10 905 (4.4)	3 686 (1.5)	1 371 (0.6)	228 (0.1)

Key : a. CWO : Correct recognition w/o rejection marks

b. CW : Correct recognition w/ rejection marks

c. IW : Incorrect recognition w/ rejection marks

d. IWO : Incorrect recognition w/o rejection marks

e. NWO : No recognition w/o rejection marks

Note : Figures in parentheses show the percentages of recognition to each total.

26. The results of scanning 10,000 pieces of questionnaires revealed that overall reading time

including waiting time for data transferring from the scanner to PC was 48.7 minutes. There were no double-feeds and jams.

27. Since Korean alphabet characters were shown to be unacceptable in recognition rate(60.4%) at the 1997 pre-test, the questionnaires were designed only with numerical characters and marks at the 1998 pre-test. The correct recognition without rejection marks was 90.4% in numerals and 99.1% in marks. Incorrect recognition without rejection marks, which are fatal errors, was 0.5% in numerals and 0.6% in marks respectively. The errors due to writing outside the given area was 0.1% in both numerals and marks. The correct recognition rates without rejection marks for households in metropolitan cities and in apartments are slightly higher than those for other types of households. There is no significant differences between short and long forms.

28. The emphasis ought to be placed on the number of questionnaires carrying recognition errors rather than on the recognition rate itself. The higher level of recognition rates does not mean less workloads in proportion; it should be huge amount of workloads for verifying recognition if the errors would be appeared on every single questionnaire.

29. However, increasing recognition rates can not be ruled out from the preparation for successful data capturing in the 2000 Census. The provision of clear number-writing instruction to respondents and visual inspection of questionnaires before being fed into the scanner would increase the recognition rates considerably.

## **. Conclusion**

30. The introduction of delivery/collect enumeration method to the 2000 Population and Housing Census in Korea is likely to be selective by the subsets of population. Due mainly to the large proportion of elderly, the rural areas turned out to be mostly lacking for the ability of completing the census questionnaire, Therefore, the application of delivery/collect method is suggested to be limited to households in urban areas. By type of housing units, apartment households seem to be more applicable to the delivery/collect method than any other types of households. Apartments have on the whole similar housing conditions, which enables enumerators to control the quality of replies. The convenience of questionnaire delivery and collection in the apartment households is another merits for the delivery/collect method from the census-operation's point of view. The costs of census-taking is expected to be reduced, to the extent not negligible, by such changes in enumeration method.

31. As far as the introduction of OCR data capture to the 2000 Census in Korea is concerned, the biggest barrier is the lack of confidence in the operational feasibility among census officials. According to the results of pre-tests, the unexpected but one of most serious problems regarding OCR data capture is the workload of eye-checking every numerical character and mark on computer monitors in the correction station. To be safe, the verification of recognition is recommended on full-scale in Korean context, irrespective of the level of recognition rates.

32. Besides the introduction of delivery/collect enumeration method and image data capture, other means which can make contributions to the cost-effective census-taking are also under study in Korea; first is to replace print media with computer media, as far as possible, which are even cheaper in data dissemination; second is to reduce the sample size

for long-form questionnaires; and, third is to utilize voluntary enumerators, to extent possible.

33. Each medium of data dissemination of the census has its own strength and weakness. Paper media are traditionally the major means for data dissemination but they are quite expensive. Computer media become preferable more and more among computer users since more data can be provided at cheaper costs and data manipulation is convenient for further analysis. The results of the 1995 Census are provided on CD-ROM in which data are stored in spreadsheet forms to facilitate data retrieval and manipulation. A single CD-ROM contains 23 books in 8 volumes of census publications. Its price is only 25,000 *won* as cheaper as one-thirteenth of the price of printed publications(320,100 *won*). For the printed publications of 1995 Census, around 333 million *won* was spent. Considering the costs of 30 million *won* spent for the production of CD-ROM in the 1995 Census, some replacement of publication with CD-ROM would reduce total census budgets.

34. From the viewpoints of cost-benefit, it is, also, necessary to evaluate whether the sample size for the long-forms in the census is efficient or not. Korean census has adopted sampling techniques in the consideration of costs and precision since 1960. Some detailed information such as labor force, child ever born, residence prior to fixed time and so forth, were enumerated on sample basis by using the long-form questionnaire. Sample size has mostly been 10% except 20% in 1960 and 5% in 1975. The method of sampling is to select every tenth EDs rather than tenth households due mainly to managerial reasons nevertheless the existence of clustering effects to increase sampling variance. Together with the homogeneity of Korean society characterized by one race, one language with small land area, the advanced sampling techniques would justify the smaller sampling size. If sampling size could be reduced, the census-taking costs could be curtailed through employing less enumerators.

35. Another approaches to the cost-effectiveness of census is to utilize voluntary enumerators. Around 1.9 million volunteers accounting for 5.4% of total population aged 15 and over are estimated to participate in voluntary activities in various fields. In particular, Local Volunteer Centers run by local governments in 124 *Shis*, *Guns* and *Gus*(*Shis* are equivalent to cities and *Ups* to towns while *Gus* are sub-divisions within Metropolitan cities) are drawing attentions of census officials. Let alone the availability of around 375 thousands persons of volunteer enrolled, the nationwide network of centers are appropriate to organize and control enumerators in census-taking. In addition, around 50 thousands of college students enrolled in the courses related to voluntary activities all over the country are considered to be another major source of census enumerators. Since daily pays for volunteers are generally one-third of those for regular census enumerators, the more the volunteers could be utilized as census enumerators, the less the costs of census-taking would be expected. The operational feasibility is going to be examined in the next pre-test scheduled in October this year.

36. The on-going study on the introduction of delivery/collect enumeration method and OCR data capture does not necessarily mean that there are no alternatives to them. On the contrary, the alternatives are simultaneously examined. For instances, the key-entry method on outsourcing basis in data capture is strongly appealing to census officials. It makes sense that the timely, accurate and cheap data capture could come true as long as the work for keying is decentralized and assisted by computer programs.

37. Making decisions on the fundamental changes in census operation should be reached as early as possible because it needs various steps in preparation of the 2000 Census. The NSO of Korea is planning to make decisions on this matter not later than at the end of this year.

## REFERENCE

ESCAP, "The 2000 Decade of Population and Housing Censuses: Recent Developments", Item 7 of the Provisional Agenda, Working Group of Statistical Experts, Tenth Session, 11-14 November 1997, Bangkok

Min-Kyung Kim, "Changes in Census Environments and Preparations for the Last Population and Housing Census in the 20th Century in Republic of Korea", A Paper Presented at the 16th Population Census Conference, 17-20 January, 1995, New Delhi, India

Min-Kyung Kim and Sam-Sik Lee, "Evaluation on Effectiveness of Introduction of OMR Technology to the 1990 Population and Housing Census", A Paper Presented at the 15th Population and Housing Census Conference, August 17-20, 1993, Seoul, Republic of Korea

NSO, Korea, 1960, 1966, 1970, 1975, 1980, 1985, 1990, 1995 *Population and Housing Census Reports*

Sam-Sik Lee. "Review of the 1995 Population and Housing Censuses and Strategies For the 2000 Population and Housing Census", A Paper presented at the 17th Population Census Conference, 4-8 November 1996, Jakarta, Indonesia

U.N. 1998. *Principles and Recommendations for Population and Housing Censuses*. Statistical Papers, No. 67/Rev. 1.

\_\_\_\_\_. 1992. *Handbook of Population and Housing Censuses, Part : Planning, Organization and Administration of Population and Housing Censuses*. Studies in Methods, No. 54.

Yeun-chung Yu. "The 1990 and 2000 World Population and Housing Censuses", A Paper Presented at the 16th Population Census Conference, 17-20 January, 1995, New

Deli, India