

OCCUPATION CODING: DO'S AND DONT'S

Version 2, August 2 2008

Harry B.G. Ganzeboom¹

Department of Social Research Methodology
Free University Amsterdam
E-mail: HBG.Ganzeboom@fsw.vu.nl

Contents

1. Choosing the occupational classification
2. Creating a coding file
3. Semi-automatic and automatic coding
4. Selecting and training coders
5. Multiple coders and dividing up the work
6. Coding
7. Quality checks
8. Archiving

Appendices:

- A. ISCO-88 major and sub-major groups
- B. Format of a coding file in long format.
- C. SPSS syntax:
 1. From wide format to long format
 2. Matching new titles with a source of already coded titles
 3. From long format to wide format
- D. Some notorious problems in coding ISCO-88
- E. MTMM model for multiple coders

¹ An earlier version of these notes was presented before the General Assembly of the International Social Survey Programme, Chicago, April 30 2008 and was further developed for a Workshop on Occupation Coding at the Academia Sinica, Taipei, July 2008. I gratefully acknowledge useful discussions with Cinzia Meraviglia, Willem Saris and Tony Tam.

1. CHOOSING THE OCCUPATIONAL CLASSIFICATION

- Purchase an ISCO-88 manual (in English) and have your coders use it. Apart from the useful Introduction (11 pages) and the classification itself (10 pages) it contains over 400 pages of definitions, examples, indexes and guidance that are useful to coders.
- The full ISCO-88 manual is also available online at the ILO website:
<http://www.ilo.org/public/english/bureau/stat/isco/isco88/>
- A frequently used alternative to official ISCO is ISCO(COM), which was created for the European Union (then: Community) by Peter Elias and associates at Warwick University. ISCO(COM) deviates from official ISCO only in minor ways (and there is nothing European about the deviations). However, the guide for users (<http://www.warwick.ac.uk/ier/isco/intro.html>) contains a lot of useful clarifying notes, some of which are reproduced here in Appendix D.
- If an ISCO translation exists in your language, use it, but be critical about details that may get lost in translation.
- Many countries have indigenous detailed occupational classifications. It is acceptable to use this as a coding frame and transfer the codes later into ISCO. (This is the procedure I use myself and I find it very helpful to review the coding by looking at the ISCO after conversion.)

2. CODING FILE

- Retype all the narrative information from the questionnaire into a database. Recommended string length: A100. Retyping should be as verbatim as possible, but it is helpful to cut some useless information (“My father was a bookkeeper”) or have it moved towards the end of the string. Spelling errors can also be corrected. Information that can be alphabetically sorted is vital to the speed of the coding process.
- Retyping occupational information into a database is low-level clerical work, but of course errors can be made. If you use Excel, be aware of the dangers of its powerful string-completion facilities, which can lead to many errors.
- Transfer all the string information into a coding file that has a ‘long’ format (see Appendix D, i.e. stack the information on respondent’s occupation, spouse’s, father’s, mother’s occupation, etc. At a minimum the coding files will contain:
 - CaseID
 - Varname
 - Occupational title (A100).
- Add other information on the occupation, preferably in numerical form, such as:
 - Industry
 - Employment status
 - Supervising status
 - Firm / farm size
 - Required qualifications (if asked in the interview!)

- DO NOT add information to the coding file other things than the occupation. In particular DO NOT add information on:
 - Education
 - Income
 - Age
 - Gender.

Coders should not be allowed to peek at education to determine an occupation code!!

- If there was double measurement of occupation (e.g. by also asking a crude question, or in longitudinal surveys from earlier waves), DO NOT include this into the coding file. Independence of measurement errors is crucial for modelling and removing it. If you are coding multiple occupations from a panel file, enter these as separate records and keep the unit ID in a separate file.
- Make sure that you have all the occupations before you start coding. Adding in information at a later stage is usually a lot of work.

3. SEMI-AUTOMATIC AND AUTOMATIC CODING

- If you have coded occupations before and kept the results, it is fine to do semi-automatic coding by matching the new information with the old information. I maintain a source of around 50.000 previously coded titles and do exact matching on the first 50 characters of the string and transfer the old codes to the new coding file. This gets me around 40-50% of the new codes, but all of these need to be reviewed since many of these are wrong or imprecise. Appendix C give a short spss syntax on how to make these matches.
- Increasingly the pro's (i.e. official statistical agencies) use knowledge databases to do in-field coding, i.e. providing interviewers and/or respondents with a potential occupation codes to choose from. Of course, such programs would be helpful to coders too. I have no experience here.

4. SELECTING AND TRAINING CODERS

- Recruitment of coders. Required qualities are: (A) Patience, (B) Interest in the world of work, (C) Willingness to look up information (library and internet skills) and correct initial judgments, (D) Ability to handle (sorting, searching, saving) the coding file, (E) honesty. In addition, it is helpful when coders are able to understand the (English language) ISCO-manual. Students can be a good choice, in particular when you know them personally and when there is chance they will encounter you again.
- Train the coders to understand ISCO-88, in particular its hierarchical digit structure. Looking at a form like Appendix A that shows the structure of ISCO using only the first two digits, is crucial. DO NOT train by showing them difficult examples.
- If you use multiple coders (DO!), it is fine that they have training sessions together. However, DO NOT allow them to consult during coding. It is crucial to have independent information.

5. MULTIPLE CODERS AND DIVIDING UP THE WORK

- It is a good idea to use multiple coders and not only because this makes it possible to speed up the work by division of labour. If coders make errors – and they always do – these will become more like random errors with multiple coders, but they may become systematic error if you employ only one coder. However, be aware that you can turn this advantage into a disadvantage if they consult each other frequently, and you deprive yourself of the opportunity to check for coder error, be it random or systematic. So DO NOT let them work together.
- Dividing up the work: If you employ multiple coders to divide up the work, give each of them a random part of the coding file. DO NOT give one coder father's and mother's occupation to code and another coder respondent's and spouse's occupation. DO NOT give one coder all occupations starting with A..M and another one N..Z.
- It is a good idea to have the random parts of the coding file partly overlap. The double coded occupations will give information on the quality of the coders.

6. CODING

- Have your coders first code the first two digits (i.e. 10 major and 28 sub-major groups). This information fits on single sheet of A4 paper (see Appendix A) – at this stage there is NO NEED to consult the manual elaborately, which is a major time-saver. This works best if the first two digits are followed by two trailing zeroes: 1100, 1200 etc.
- Frequently sorting the coding file is crucial for quick and accurate coding. So your coders should be able to sort the coding file themselves. Sorting by title helps much to group similar titles. Sorting by code helps much to check the results. [NB: if you use Excel, be aware of the pitfalls of sorting in this program that can sort columns separately!]
- If your coders have done the coding by the first two digits, this might be a natural point to review and discuss their results with them. Make sure you instruct them, DO NOT correct them. Then they can start filling in the details for minor and unit groups (3rd and 4th digit), using the manual. Tell them that it is not a problem changing the first two digits at this stage, if they think this is an improvement. At this stage it is most efficient to work from a file that is sorted by code and within codes by title, and to consult the manual frequently.
- Encourage your coders, in particular as they are confronted with vague and difficult information. Tell them this is not their fault, but yours when you designed the questionnaire and/or instructed the interviewers. Any occupation coding ends up with 10%-20% ambiguous pieces of information for which they have to make a best guess.
- An experienced and skilful coder can process on average around 250 occupations per hour.

7. CHECKING THE QUALITY

- Add ISCO-88 value labels to the codes to see whether coders have used illegitimate codes. These labels are available on: <http://home.fsw.vu.nl/hbg.ganzeboom/isko88/index.htm>.

- Check the quality of the coders by reconstructing the coding file into a person file ('wide format') and calculate correlations between occupations (father, mother, respondent, spouse) after scoring the codes with a status scale (such as ISEI). The coder who obtains the highest correlation is the best, or at least the most consistent one. The one(s) with the lower correlations need(s) further scrutiny.
- If you have employed multiple coders who have done double coding, a MTMM model (see Figure E2) will allow you to determine (A) coder reliability (coefficients **a** and **b**), (B) coder unique systematic bias (coefficients **e** and **f**), (C) the true score correlation (coefficient **d**), (D) construct-reliability (coefficient **c**). You only need a correlation matrix and some high-school algebra to estimate this model. The coefficients of most interest in this context are **a** and **b**. It is to be expected that **e** and **f** are essentially zero. If they are not, you need to look at systematic differences between coders, which may be restricted to the systematic misclassification of a few numerous occupations (e.g. farmers).
- If you have used double coding and you are not a lisrelite, you need to harmonize the codes into a single one, using an adjudicator (this may be yourself). Make sure that you do not do this until all the independent double coding has been done. You can use the results of the MTMM model (a versus b) to give preference of the results of one coder above the other. Also make sure that you DO NOT change the original codes, but indeed produce a third one. If you are a lisrelite, the whole adjudication procedure is unnecessary and the original two codes are more valuable. DO NOT adjudicate with the initial coders being present, be independent!

8. ARCHIVING

- The coding file should be kept and archived with the original data file, for the following reasons:
 - Can be used as a coding frame at future occasions.
 - Can be used to upgrade or convert information to other or new classifications in the future. ISCO-08 is coming up!
 - It can be analysed by lisrelites. Coding difference may add between 5%-10% attenuation of structural coefficients and this can be corrected by using a multiple indicator model.

Appendix A: ISCO-88 Major and Sub-Major Groups

1000 LEGISLATORS, SENIOR OFFICIALS & MANAGERS

- 1100 LEGISLATORS & SENIOR OFFICIALS
- 1200 CORPORATE MANAGERS [LARGE ENTERPRISES]
- 1300 [SMALL ENTERPRISE] GENERAL MANAGERS

2000 PROFESSIONALS

- 2100 PHYSICAL, MATHEMATICAL & ENGINEERING SCIENCE PROFESSIONALS
- 2200 LIFE SCIENCE & HEALTH PROFESSIONALS
- 2300 TEACHING PROFESSIONALS
- 2400 OTHER PROFESSIONALS

3000 TECHNICIANS AND ASSOCIATE PROFESSIONALS

- 3100 PHYSICAL & ENGINEERING SCIENCE ASSOCIATE PROFESSIONALS
- 3200 LIFE SCIENCE & HEALTH ASSOCIATE PROFESSIONALS
- 3300 TEACHING ASSOCIATE PROFESSIONALS
- 3400 OTHER ASSOCIATE PROFESSIONALS

4000 CLERKS

- 4100 OFFICE CLERKS
- 4200 CUSTOMER SERVICES CLERKS

5000 SERVICE WORKERS & SHOP & MARKET SALES WORKERS

- 5100 PERSONAL & PROTECTIVE SERVICES WORKERS
- 5200 [SALESPERSONS, MODELS & DEMONSTRATORS]

6000 SKILLED AGRICULTURAL & FISHERY WORKERS

- 6100 MARKET-ORIENTED SKILLED AGRICULTURAL & FISHERY WORKERS
- 6200 SUBSISTENCE AGRICULTURAL & FISHERY WORKERS

7000 CRAFT ETC TRADES WORKERS

- 7100 EXTRACTION & BUILDING TRADES WORKERS
- 7200 METAL, MACHINERY ETC TRADES WORKERS
- 7300 PRECISION, HANDICRAFT, PRINTING ETC TRADES WORKERS
- 7400 OTHER CRAFT ETC TRADES WORKERS

8000 PLANT & MACHINE OPERATORS & ASSEMBLERS

- 8100 STATIONARY-PLANT ETC OPERATORS
- 8200 MACHINE OPERATORS & ASSEMBLERS
- 8300 DRIVERS & MOBILE-PLANT OPERATORS
- 8400 SEMI-SKILLED WORKERS NFS

9000 ELEMENTARY OCCUPATIONS

- 9100 SALES & SERVICES ELEMENTARY OCCUPATIONS
- 9200 AGRICULTURAL, FISHERY ETC LABOURERS
- 9300 LABOURERS IN MINING, CONSTRUCTION, MANUFACTURING & TRANSPORT

Appendix B: The Italian coding file

Variable Information

Variable	Position	Label	Print Format
INTNR	1	Interview number	F8
VARNAME	3	Father / Respondent	A8
TIPO_LAV	4	Job Title	A200
DES_LAV	5	Description	A200
POS_OCC	6	Status in Occupation	A54
SETTORE	7	Industry	A90
ISKO3	8	Fabiana	F8
ISKO4	9	Federica	F8
ISKO34	10	Bestcode	F8

Tabel B1: Part of the Italian coding file, unsorted

VARNAME	TIPO_LAV	DES_LAV	POS_OCC	SETTOE	ISK03	ISK04
isko	impresa di pulizie	puliva uffici della dogana	artigiano	altri servizi alle imprese	9132	9132
flisko	imbiachino			altri servizi al consumatore finale	7141	7141
isko	ristorante	aiutante cuoco	contratto stagionale	ristoranti, bar e cantine	5122	5122
flisko	costruzioni			costruzioni	1210	7122
isko	clinica privata	inserviente	contratto di altro tipo	altri servizi alle imprese	5132	5132
flisko	barca a rema			trasporti marittimi e per vie d'acqua	8340	8340
isko	imbarcazione	operaio motorista	lavoro a chiamata	trasporti marittimi e per vie d'acqua	9312	8340
flisko	costruzioni			costruzioni	1210	7122
isko	costruzione	autista	contratto di manodopera in leasing	costruzioni	1210	8330
flisko	operaio muratore impresa edile			costruzioni	7129	7122
isko	salariato agricolo	trattorista	contratto di manodopera in leasing	agricoltura, caccia e relativi servizi	9211	8331
flisko	tornitore meccanico			manutenzione e riparazione autoveicoli e motoveicoli, vendita componenti e accessori	7223	7223
isko	stiratrice	stiro per conto di persone di volta in volta mi contattano	lavoro a chiamata	79	8264	9133
flisko	contadino			agricoltura, caccia e relativi servizi	6111	6111
isko	collaboratrice domestica	mi occupavo di fare i servizi	lavoro a chiamata	77	9131	9131
flisko	operaio			costruzioni	7122	9313
isko	parrucchiere	parrucchiere	artigiano	artigiano	5141	5141
flisko	muratore, lavorava con lo zio era una piccola impresa			costruzioni	7122	7122
isko	lavoro come operaio in una azienda che produce gomma plastica, teloni in pvc, tende da sole	operaio	contratto di manodopera in leasing	fabbricazione articoli in gomma e materie plastiche	8232	8232
flisko	operaio azienda privata			calzaturificio	8266	8266
isko	corrispondenza estera e traduzioni in una casa editrice		contratto di manodopera in leasing	settore grafico editoriale	2444	2444
isko	lavoro contabile o durante le fiere gli davvo una mano		libero professionista	non ricorda	3433	4190
flisko	autista di camion			costruzioni	8324	8324
isko	autista di camion	guida camion, pale meccaniche	contratto di manodopera in leasing	costruzioni	8324	8330
flisko	commerciante di mobili			vendita al dettaglio di altri beni nuovi	1314	1314
isko	bracciante agricola	raccolta olive	contratto stagionale	agricoltura, caccia e relativi servizi	9211	9211
flisko	magazziniere			vendita al dettaglio di altri beni nuovi	4131	4131

Tabel B2: Part of the Italian coding file, sorted by TIPO_LAV

VARNAME	TIPO_LAV	DES_LAV	POS_OCC	SETTOE	ISKO3	ISKO4
flisko	autista di camion			costruzioni	8324	8324
isko	autista di camion	guida camion, pale meccaniche	contratto di manodopera in leasing	costruzioni	8324	8330
flisko	barca a rema			trasporti marittimi e per vie d'acqua	8340	8340
isko	bracciante agricola	raccolta olive	contratto stagionale	agricoltura, caccia e relativi servizi	9211	9211
isko	clinica privata	inserviente	contratto di altro tipo	altri servizi alle imprese	5132	5132
isko	collaboratrice domestica	mi occupavo di fare i servizi	lavoro a chiamata	77	9131	9131
flisko	commerciante di mobili			vendita al dettaglio di altri beni nuovi	1314	1314
flisko	contadino			agricoltura, caccia e relativi servizi	6111	6111
isko	corrispondenza estera e traduzioni in una casa editrice		contratto di manodopera in leasing	settore grafico editoriale	2444	2444
isko	costruzione	autista	contratto di manodopera in leasing	costruzioni	1210	8330
flisko	costruzioni			costruzioni	1210	7122
flisko	costruzioni			costruzioni	1210	7122
isko	imbarcazione	operaio motorista	lavoro a chiamata	trasporti marittimi e per vie d'acqua	9312	8340
flisko	imbiachino			altri servizi al consumatore finale	7141	7141
isko	impresa di pulizie	puliva uffici della dogana	artigiano	altri servizi alle imprese	9132	9132
isko	lavoro contabile o durante le fiere gli davo una mano		libero professionista	non ricorda	3433	4190
isko	lavoro come operaio in una azienda che produce gomma plastica, teloni in pvc, tende da sole	operaio	contratto di manodopera in leasing	fabbricazione articoli in gomma e materie plastiche	8232	8232
flisko	magazziniere			vendita al dettaglio di altri beni nuovi	4131	4131
flisko	muratore, lavorava con lo zio era una piccola impresa			costruzioni	7122	7122
flisko	operaio			costruzioni	7122	9313
flisko	operaio muratore impresa edile			costruzioni	7129	7122
flisko	operaio azienda privata			calzaturificio	8266	8266
isko	parrucchiere	parrucchiere	artigiano	artigiano	5141	5141
isko	ristorante	aiutante cuoco	contratto stagionale	ristoranti, bar e cantine	5122	5122
isko	salariato agricolo	trattorista	contratto di manodopera in leasing	agricoltura, caccia e relativi servizi	9211	8331
isko	stiratrice	stiro per conto di persone di volta in volta mi contattano	lavoro a chiamata	79	8264	9133
flisko	tornitore meccanico			manutenzione e riparazione autoveicoli e motoveicoli, vendita componenti e accessori	7223	7223

Tabel B3: Part of the Italian coding file, sorted by ISKO4

VARNAME	TIPO_LAV	DES_LAV	POS_OCC	SETTOE	ISKO3	ISKO4
flisko	commerciante di mobili			vendita al dettaglio di altri beni nuovi	1314	1314
isko	corrispondenza estera e traduzioni in una casa editrice		contratto di manodopera in leasing	settore grafico editoriale	2444	2444
flisko	magazziniere			vendita al dettaglio di altri beni nuovi	4131	4131
isko	lavoro contabile o durante le fiere gli davo una mano		libero professionista	non ricorda	3433	4190
isko	ristorante	aiutante cuoco	contratto stagionale	ristoranti, bar e cantine	5122	5122
isko	clinica privata	inserviente	contratto di altro tipo	altri servizi alle imprese	5132	5132
isko	parrucchiere	parrucchiere	artigiano	artigiano	5141	5141
flisko	contadino			agricoltura, caccia e relativi servizi	6111	6111
flisko	costruzioni			costruzioni	1210	7122
flisko	costruzioni			costruzioni	1210	7122
flisko	operaio muratore impresa edile			costruzioni	7129	7122
flisko	muratore, lavorava con lo zio era una piccola impresa			costruzioni	7122	7122
flisko	imbiachino			altri servizi al consumatore finale	7141	7141
flisko	tornitore meccanico			manutenzione e riparazione autoveicoli e motoveicoli, vendita componenti e accessori	7223	7223
isko	lavoro come operaio in una azienda che produce gomma plastica, teloni in pvc, tende da sole	operaio	contratto di manodopera in leasing	fabbricazione articoli in gomma e materie plastiche	8232	8232
flisko	operaio azienda privata			calzaturificio	8266	8266
flisko	autista di camion			costruzioni	8324	8324
isko	costruzione	autista	contratto di manodopera in leasing	costruzioni	1210	8330
isko	autista di camion	guida camion, pale meccaniche	contratto di manodopera in leasing	costruzioni	8324	8330
isko	salariato agricolo	trattorista	contratto di manodopera in leasing	agricoltura, caccia e relativi servizi	9211	8331
flisko	barca a rema			trasporti marittimi e per vie d'acqua	8340	8340
isko	imbarcazione	operaio motorista	lavoro a chiamata	trasporti marittimi e per vie d'acqua	9312	8340
isko	collaboratrice domestica	mi occupavo di fare i servizi	lavoro a chiamata	77	9131	9131
isko	impresa di pulizie	puliva uffici della dogana	artigiano	altri servizi alle imprese	9132	9132
isko	stiratrice	stiro per conto di persone di volta in volta mi contattano	lavoro a chiamata	79	8264	9133
isko	bracciante agricola	raccolta olive	contratto stagionale	agricoltura, caccia e relativi servizi	9211	9211
flisko	operaio			costruzioni	7122	9313

Appendix C: SPSS syntaxes

From wide to long format.

```
** OCCUPATION STRINGS ARE LOCATED IN FOCC ROCC

string title (A100) / varname (A8).

compute title=substr(focc,1,100).
compute varname="FATH".
save outfile="fff.sav"/keep=respnr varname title.

compute title=substr(rocc,1,100).
compute varname="RESP".
save outfile="rrr.sav"/keep=respnr varname title.

add files /file="fff.sav" /file="rrr.sav".

sort cases by title.
```

Matching with source file that has coded occupations

```
** MATCHING WITH SOURCE FILE THAT HAS CODED OCCUPATIONS **.
** VARIABLES OF THE SOURCE FILE ARE: TITLE, ISKO.

get file="codingfile.sav".

string key (a50).

comp key=substr(title,1,50).
sort cases by key.

add files /file="source.sav" /file=* in=new.

recode new (sysmiss=0).

do if (new eq 0).
comp key=substr(title,1,50).
end if.

sort cases by key new.

do if (sysmis(isko)).
if (lag(key) eq key) isko=lag(isko).
end if.

temp.
select if (new eq 1).
freq isko.

select if (new eq 1).

save outfile="codingfile_coded.sav".
```

From long to wide coding file.

```
get file="codingfile_coded.sav".

if (varname eq "FATH") fisko=isko.
if (varname eq "RESP") risiko=isko.

temp.
select if (varname eq "FATH").
save outfile="fff_coded.sav" /keep=respnr fisko.

temp.
select if (varname eq "RESP").
save outfile="rrr_coded.sav" /keep=respnr risiko.

match files /file="fff_coded.sav" /file="rrr_coded.sav" /by respnr.

define @isko()
  fisko risiko
!enddefine.

define @isei()
  fisei risei
!enddefine.

include "iskoisei.sps".

corr fisei risei.
```

Appendix D: SOME NOTORIOUS PROBLEMS IN CODING ISCO-88 AND THEIR SOLUTIONS

Managers, supervisors, proprietors

- Managers: there are two sub-major groups of managers: [1200] Corporate Managers and [1300] General Managers. Corporate Managers work in large – multi-department -- firms that have at least 2 other managers. General Managers work in small firms with at most one other manager. In practice, the distinction will have to be made by title and/or firm size.
- At the minor group level, there are two kinds of department managers: [1220] Production and Operations Department Managers and [1230] Other Department Managers. So, [1220] manage department that do the core business, [1230] head support departments. One will often have to look at industry to make the distinction.
- Working proprietors [small firm] go into general managers.
- Work supervisors [Foreman] go into [1200], if supervising is their only task; they go into [1300] if it is their dominant task, but when they still work along with their subordinates, and they are coded with their subordinates if they are only a ‘lead worker’.

Professionals and associate professionals in teaching and nursing

[3230] Associate Professionals in Nursing and [3300] Associate Professionals on Teaching are associate [=assistant] to [2230] Nursing Professionals, and [2330]-[2340] Teaching Professionals. The classification assumes that they are at least two levels in these fields of work and leaves it to the coders to make the distinction. If there is no such differentiation in your country, choose the highest (professional) alternative.

Note that primary and pre-primary teachers always have to be coded as Professionals, even if you feel that they are not at the same skill level as, say, university professors,

Farmers and farm workers

Farmers and farm workers can be coded in 5 places in ISCO-88

[1210] Production Department Manager in Agriculture

[1311] General Managers in Agriculture

[6100] Market Oriented Skilled Agricultural Workers

[6200] Subsistence Agricultural Workers

[9200] Agricultural Labourers

[1210] would be rare – it can only occur in very large farms that have other (support) departments. The choice between [6100] and [1311] is hard to make, and the ISCO manual does not help very much. Most often self-employed farmers are coded as [1311] and [6100] is left empty. However, I think that users would be better helped with coding small farmers in [6100], as [1311] is a four-digit code and easily vanishes in processing.

Craft workers and machinery workers

At the major group level, ISCO distinguishes between [7000] Craft-Trades Workers and [8000] Machine Operators. While related, the distinction is not the same as between skilled

workers and semi-skilled workers. Coding problems arise because many related occupations have parallels in the two major groups, e.g.:

[7110] and [8110]	Miners
[7340] and [8250]	Printers
[7410] and [8270]	Food Producers
[7420] and [8140]	Wood Treaters
[7430] and [8260]	Textile Producers

Mostly, interview materials do not give very much of a clue whether someone is/was a machine worker or not. If no further evidence is available, I advice to use the [7000] variant.

Crude descriptions

Very often respondents will give only crude descriptions of occupations, which is not saying that these descriptions are unclear. The major solution is to code these descriptions on the 1- or 2-digit level (and use trailing zeroes). Some examples:

Managers	[1200]
Shop Owner	[1300]
Independent	[1300]
Entrepreneur	[1210]
Foreman	[1319]
Skilled Worker	[7000]
Semi-Skilled Worker	[8000]
Unskilled Worker	[9000]

In practice, this works quite well, in particular when one is primarily interested in occupational status.

Ambiguous and multiple descriptions

Very often respondents give information that can be interpreted in multiple ways, because a job has multiple components, or someone has multiple jobs. The Introduction to the ISCO provides a set of rules that can be adapted to this situation. Code the information using the following rules in sequence:

- **Numerical dominance rule:** when one activity dominates, or one interpretation is the far more plausible one given the distribution of activities in the population, code accordingly,
- **Skill level rule:** if a set of activities involves a mix of skill levels, choose the most skilled one.
- **Production rule:** if a set of activities involves production next to sales and/or management, choose the production occupation.

A list of important minor and unit groups distinctions

Not all distinctions between minor (3-digit) and unit (4-digit) groups are equally relevant in a sociological sense, as (sub-)major groups are fairly homogeneous in status. However, there are important exceptions to this rule, and these are worth special checking when reviewing the codes:

[1221] [1311]	Farmers are very different from other Managers
[2220] [2230]	Doctors versus Nurses
[2310] – [2390]	Various levels of Teachers

However, shortness of this list should not prevent you from using the full four-digit code. Adding in the 3rd and 4th digit is actually not so much work, once you have done the first two.

Appendix E: MTMM models for multiple coders

If you employ multiple coders (to divide up and speed up the coding process) you can check the quality of their work by having them overlap for a piece of for the entire coding file. You can learn about the quality of their work not only by comparing their choice of the overlapping occupations, but if you have multiple occupations to code (e.g. respondent and spouse), you can also compare them using the between-occupations correlation, provided that you have divided up the coding file in random parts. I recommend to create overlap next to randomised separation of the coding file. Independent of the size of the coding file, the overlap should be at least 200-400 (XX?) randomly chosen occupations to get to sufficiently small confidence intervals.

I present two examples where 2 coders were used for father-respondent data, one for Italy² (2 coders coded 1800 pairs, and one for Taiwan, where two groups³ of coders codes 375 pairs. In both cases there was complete overlap of the occupations between the coders. After transferring the coding file into a wide format and scoring the ISCO codes by the ISEI metric, we obtain the following correlation matrices.

Figure E1: Between-coder correlations in two coders – two occupations situation, two countries								
	Italy, 2005, N=1800 occupations				Taiwan, 2007, N=375 occupations			
	Fisei1	Fisei2	Risei1	Risei2	Fisei1	Fisei2	Risei1	Risei2
Fisei1	1.000				1.000			
Fisei2	0.772	1.000			0.865	1.000		
Risei1	0.352	0.332	1.000		0.389	0.369	1.000	
Risei2	0.321	0.322	0.811	1.000	0.448	0.442	0.821	1.000

First we look at the correlations that directly measure the between-coder agreement. These correlations should be high and they are. On detailed inspection, we can make the following further observations:

- Taiwanese coders are on average a bit better than the Italians; or: Taiwanese occupations are a bit easier to code than Italian.
- The Italians are slightly more in agreement about respondents than about fathers, but for the Taiwanese this is the other way around. In my experience the Taiwanese pattern is the more common one.
- No one is perfect, coders introduce considerable error in the measurement process. Otherwise, there correlations would be 1.00. Note that the average between-coder correlation for the same occupation (say, 0.81) suggests that for each statistical relationship with occupation, 10% gets lost ('attenuated') by coding error alone!

² The data were kindly provided by Cinzia Meraviglia. The data were collected in 2005 by Luca Ricolfi and the coders were Fabiana and Federica.

³ One group consisted of the local fieldwork supervisors who initially coded the data in-field, the other group were 49 attendants to a Workshop in Occupation Coding in Taipei, each of whom coded 25 occupations. I averaged the code of men and women.

Note that we do not learn anything yet about the relative quality of the two coders at this stage. However, we can learn about that by inspecting the four intergenerational correlations. These correlations are produced by three components:

1. The true intergenerational correlation between fathers and respondents. This does not differ between coders, and is not important for evaluating the differences between them, but it is often what we are most interested in.
2. Random error brought in by coders. If one coder brings in more random error than the other, this will decrease both his own intergenerational correlation, and the intergenerational correlation with the other coder.
3. Systematic (correlated) error, i.e. ways in which coders systematically (i.e. identical for fathers and respondents) misclassify occupations. This may increase the within-coder intergenerational correlation, but not the between-coder one (unless they have brought in the same biases, which is indistinguishable from the true intergenerational correlation).

If there is only random error and this is the same for all coders, the four intergenerational correlations will be identical (but they may all be downwardly biased). This is pretty much the case in both matrices, so there are only minor differences between coders in random error and their systematic differences must be small.

A formal structural equation model can be used to estimate the parameters precisely (see Figure E.1) in LISREL, AMOS or similar software. However, with some high school algebra you get pretty good results using only paper and pencil, in particular if you assume that there is negligible systematic bias (effects e and f). It involves solving the following four equations:

$$c = \sqrt{(r_{12}/r_{34})}$$

$$a*b = r_{12}*c = r_{34}*1$$

$$a/b = \sqrt{(r_{13}/r_{24})}$$

$$d = r_{13}/a*a = r_{24}/b*b$$

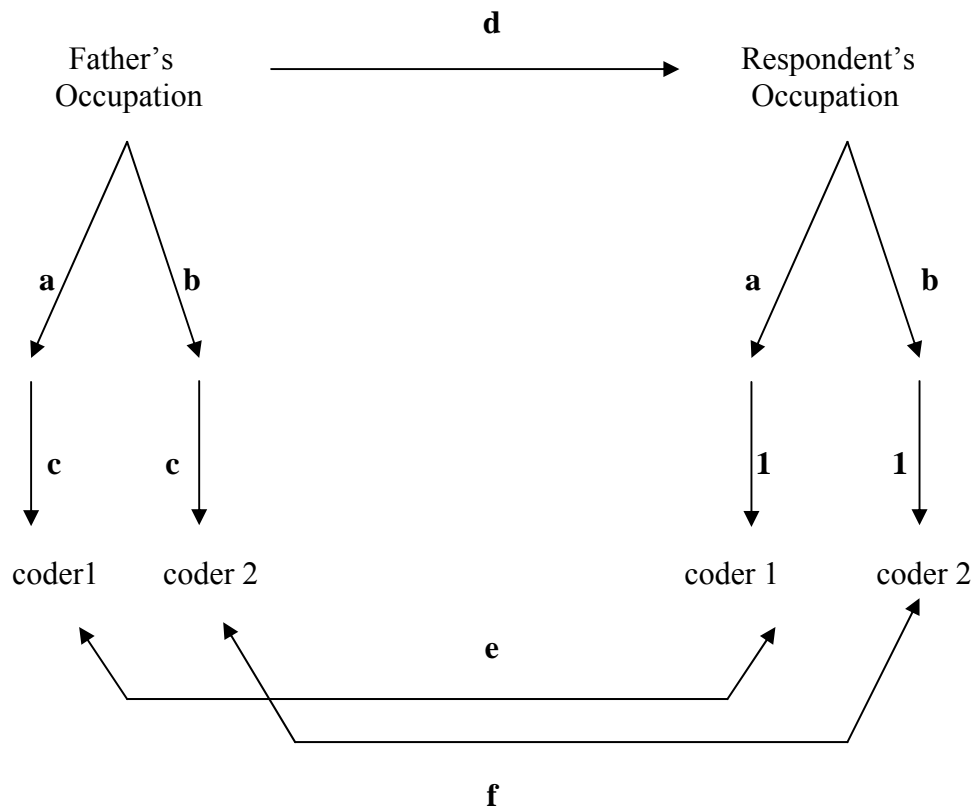
As the algebraic system is overdetermined (4 unknowns with 6 independent correlations), the result of the last step may be inconsistent. This would suggest that effect e and f matter. Unfortunately, in this system one cannot independently identify a/b relative to e/f. If e/f is of great concern to you, you should bring in more variables (e.g. father's and/or respondent's education) and you can identify these ratio's.

Using Lisrel, we obtain the following estimates

	Italy	Taiwan	
a	0.925	0.902	coder 1
b	0.877	0.954	coder 2
c	0.975	0.982	construct reliability
d	0.419	0.492	true score correlation

If you have more coders, or – even better – more occupations, it is easier to identify coding error, both random and systematic. If tested in these two examples, the differences between coders are not statistically significant.

Figure E2: A simple MTMM model to identify random and systematic coding error.



If $r_{12}/r_{34} > 1$, interchange **c** and **1**.