Online Appendix for

"Changes in Returns to Task-Specific Skills and Gender Wage Gap"

By Shintaro Yamaguchi

In the DOT, many characteristics are measured by a multi-point scale and have detailed definitions. For example, the variable DATA measures the complexity of tasks in relation to information, knowledge, and conceptions, with integers from 0 to 6. Tasks at the lowest level of complexity involve judging the readily observable characteristics of data. Examples include sorting hats according to color and size specified, comparing invoices of incoming articles with the actual number and weight of articles, and so on. Tasks at the intermediate level of complexity involve compiling information. Examples include summarizing details of transactions, collecting, classifying, and recording data, and receiving customer complaints to record and file them for future processing. Tasks at the highest level of complexity involve integrating analysis of data to discover facts and developing the knowledge of concepts for interpretations. Examples include formulating hypotheses and experimental designs, writing critical reviews of art for publication, and conducting research. Other tasks such as operating machines or equipment are also evaluated in a similar manner. Some tasks are measured by a binary variable that takes the value one if the occupation involves the task and zero otherwise.

Tables 1 and 2 show variables that measure cognitive and motor task complexity. They are highly correlated within the task group, and thus the information can be summarized by a single index using Multiple Correspondence Analysis (MCA).

Correspondence analysis is a generalized principal component analysis tailored for the analysis of qualitative data. MCA is an extension of correspondence analysis which allows one to analyze the pattern of relationships of several categorical dependent variables. In short, MCA is a dimension reduction technique for categorical variables. There are 10 categorical variables for the cognitive task complexity index and 8 categorical variables for the motor task complexity index. These variables are converted into 52 and 43 indicator variables, respectively. In MCA, variation of the data is called inertia, which is the sum of chi-square distances to the centroid. In calculating inertia, I account for off-diagonal subtables of the Burt matrix only.

I apply the MCA to the April 1971 CPS, augmented by the fourth edition of the DOT.[i] This is the only data that contains the 1970 census occupation code, the DOT occupation code, and the DOT variables, which allows for linking the DOT variables and the census occupation code. The task complexity indices constructed by MCA account for 50% and 39% of the inertia of the 1971 CPS sample, respectively. The resulting indices at the individual level are aggregated into the level of the 1970 census occupation by taking the mean for each of the 1970 census 3-digit occupations using the sampling weights so that they can be merged with the PSID. The indices are normalized so that the mean is 0.5 and the standard deviation is 0.1 for the working individuals in the 1971 augmented CPS.

Appendix Table 1

DOT Variables for Cognitive Task Complexity Index

Variable	No of Categories/Levels
Worker Function	
DATA	7
PEOPLE	9
General Educational Development	
reasoning	6
mathematics	6
language	6
Aptitude	
intelligence	4
verbal	5
numerical	5
Temperament	
influencing people	2
dealing with people	2

Appendix Table 2

DOT Variables for Motor Task Complexity Index

Variable	No of Categories/Levels
Worker Function	
THINGS	8
Aptitude	
motor coordination	5
finger dexterity	5
manual dexterity	5
eye-hand-foot coordination	5
spatial	5
form perception	5
color discrimination	5

[i] The data file is available at the ICPSR website (http://dx.doi.org/10.3886/ICPSR07845.v2).