STEM Classification in the Formal Labor Market in Brazil

May 2021

Abstract

STEM fields have gained increasing notoriety in recent years with the purpose of developing human capital and innovation across the world. However, there is no universal definition of STEM. In some countries, particularly in Brazil, the debate incipient and almost restricted to education. This paper aims to fill the gap in the literature and provide a STEM definition for occupations in Brazil using the *CBO-2002*, with full disclosure of the classification process. We also apply our definition to labor market data using *RAIS*, presenting a picture of the country's STEM market.

1 Introduction

STEM fields (Science, Technology, Engineering and Mathematics) have gained increasing notoriety, in particular due to educational policies implemented in several countries in recent years with the aim of developing human capital and innovation. There is no universal definition of STEM, but there is reasonable agreement that STEM specialists use their knowledge in these disciplines to solve problems and provide scientific and technological advances. Among developed countries like the United States, the acronym is widely used in discussions concerning academia, government and business in general, since its emphasis on innovation have strong implications for the labor market and the economy.

In Brazil, however, the STEM debate is still incipient. Not only the term is is little known but also the definition of STEM is even more vague. In order to fill this gap and provide a better understanding of what STEM is in Brazil, this document categorizes occupations in Brazil into STEM and non-STEM using the *Classificação Brasileira de Ocupações* (CBO) system, the official occupational classification of the Ministry of Labour. We also apply our classification to *Relação Anual de Informações Sociais* (RAIS)[1], an administrative data set from the Ministry of Labor in which occupations are identified by a *CBO-2002* code, to look at the share of STEM workers in the formal labor market in Brazil.

In this sense, this paper offers an unprecedented classification method using the *CBO-2002*, explaining in detail each step taken to reach what we believe to be the best choice of classification and why. This document includes other classification we attempted using different aggregation levels and different conversion methods, and describe the reasons we chose not to use them, as well as each method's limitations. All of the STEM occupations classifications proposed here were based on the same definition[2] of STEM occupations, from *United States Bureau of Labor Statistics* (*USBLS*) "Options for defining STEM occupations

[1]RAIS is a matched employer-employee data that contains the universe of formal workers and firms in Brazil. Since it includes unique identifiers for firms and individuals, it enables us to track worker's history in the labor market

[2]*Attachment C*: Detailed 2010 *SOC* occupations included in STEM. Options for defining STEM (Science,Technology, Engineering, and Mathematics) occupations under the 2010 Standard Occupational Classification (SOC) system. August 2012. Available at < https://www.bls.gov/soc/attachment_c_stem.pdf >. Accessed on: 01/03/2021

under the 2010 SOC" series of documents.

Among our contributions, we also look at the labor market data using *RAIS* to better understand the share of the labor market that each classification is accounting for. According to our classification, in 2017, Brazil had around 924 thousand of STEM workers, which represented 2.9% of the formal labor market. Compared to the United States, this is around eight times less than the 8,620 thousand of US STEM workers reported in 2015 (Noonan, 2017), which accounts for 5.7% of US workers. These numbers provide a first glance of how small is the STEM market in Brazil, in relation to developed countries.

The literature concerning the STEM labor market in Brazil is still very scarce. Few other authors have classified Brazilian occupations into STEM and non-STEM. Schwartzman (2018) looks at the educational distribution of the labor market in Brazil with emphasis on STEM jobs[3]. Using data from *PNAD*, the National Household Sample Survey that contains information on both formal and informal workers in Brazil, he finds that among workers with a post secondary education, only 10% are in the STEM fields. Similar to this paper, Bonini et al. (2020) look at data from *RAIS*, and present preliminary results on the STEM labor market in Brazil of an ongoing research. They find that STEM workers are only 0.79% of the labor force in Brazil. They also classify STEM occupations using *CBO*, but follow the *Economics and Statistics Administration* (ESA) criteria. Beyond these two mentioned contributions, the authors are unaware of other works that investigate the STEM labor market in Brazil.

Our classification is distinguished from the ones mentioned above because it directly compares the four digit *CBO-2002* occupations with the *Standard Occupational Classification* (SOC) STEM occupations in the classification recommended by the *USBLS*. As will become evident further, this method minimizes inefficiency, loss of relevant information, and severe loss of variation due to multiple changes in aggregation levels of the two classification systems. Furthermore, it provides full disclosure of the classification and coding process, allowing compatibilization with the four digit *CBO-2002*.

Besides this Introduction, the paper is organized as follows: Section 2 describes each country's occupational classification systems, and compares them, as well as *USBLS* STEM

[3]The mentioned work does not provide public access to a table with the occupations classified as STEM nor give enough details on how the classification was done, therefore not being reproducible.

definition. Section 3 outlines briefly our classification attempts, and describes carefully the classification method we adopted and found to be best. The Appendix details the classification methods we did not follow through with, and explain the reasons behind this decision. We also provide the classification table and occupational codes in the Appendix.

2 Occupational Classifications Systems

2.1 USBLS STEM Definition

To classify Brazillian occupations into STEM and non-STEM, we used the *United States Bureau of Labor Statistics* (USBLS) definition on STEM occupations as a starting point. In 2012, the USBLS *SOC Policy Committee* (SOCPC) elaborated a document with options for defining STEM occupations based on the *2010 SOC* detailed occupations, in order to enhance comparability of data across statistical agencies and organizations studying the STEM workforce for policymaking purposes (USBLS, 2012a,b,c).

The document, called *'Options for defining STEM occupations under the 2010 SOC'* is divided into three attachments. Attachment A (USBLS, 2012a) explains the classification criteria and identifies two major STEM domains: (1) *Science, Engineering, Mathematics, and Information Technology Domain*; and (2) *Science- and Engineering-Related Domain*. Each domain, in turn, is subdivided into two sub-domains, as shown in Table 1.

Table 1: USBLS reccomendation of STEM domains

Domain	Sub-domain
Science, Engineering, Mathematics, and Information Technology Domain	**1. Life and Physical Science, Engineering, Mathematics, and Information Technology Occupations**
	3. Architecture Occupations
Science and Engineering-Related Domain	2. Social Science Occupations
	4. Health Occupations

Source: *Bureau* of Labor and Statistics, 2012.

Attachment B (USBLS, 2012b) presents a grid that divides each of the 4 sub-domains into 5 types of STEM occupations, which is not relevant for this paper classification. Attachment C (USBLS, 2012c) is a table with all of the detailed *SOC-2010* occupations and codes. Each

recommended STEM occupation is marked with a color that identifies them with one or more of the above sub-domains, while the rest of the occupations are shown in white.

For our STEM classification, we considered occupations from sub-domain 1, marked as '*Life and Physical Science, Engineering, Mathematics, and Information Technology Occupations*', indicated by the color gray in *Attachment C* table. Because *SOC* occupations are classified based on the work performed, some detailed *SOC* occupations combine work activities that are appropriate to more than one STEM sub-domain. (USBLS, 2012a). In these cases, if an occupation was split across 2 sub-domains, and one of them belonged to sub-domain 1, we considered it as a STEM occupation. Occupations from the remaining sub-domains were not considered STEM.

2.2 Standard Occupational Classification (SOC)

To make the *USBLS* definition of STEM occupations compatible with the occupations from the Brazilian classification, it was first necessary to make the occupational classification systems from the two countries comparable. The US classification system, also known as the *Standard Occupational Classification* (SOC), is a classification system designed to reflect the current occupational structure of the country. It classifies all occupations in which work is performed for pay of profit, and is organized in a tiered system with four levels, ranging from major groups to detailed occupations (see Table 2).

At the most specific level, there are 840 detailed occupations, each uniquely identified by a six-digit code. Detailed occupations with similar job duties, and in some cases skills, education, and/or training are grouped together in SOC. In *SOC-2010*, "each worker is classified into only one of the 840 detailed occupations based on the tasks he or she performs" (USBLS, 2010). Although there is a new version of the classification system with some minor changes (*SOC-2018*)[4], we will be focusing on the *SOC 2010*, as *Attachment C*, the *USBLS* document we reference for dividing jobs into STEM and non-STEM, uses *SOC-2010* codes to classify occupations.

[4]In the *SOC-2018*, no substantive change occurred for about 88 percent of the detailed occupations relative to *SOC-2010*. See USBLS (2018).

Table 2: SOC-2010 Structure

Structure	# Digits	# Occupations
Major Groups	2	23
Minor Groups	3	97
Broad Occupations	4	461
Detailed Occupations	6	840

Source: *Bureau* of Labor and Statistics, 2012.

2.3 *Classificação Brasileira de Ocupações* (CBO)

Similar to the *Standard Occupational Classification*, the *Classificação Brasileira de Ocupações* (CBO), by the *Brazilian Ministry of Labor* (MTE, in portuguese), defines an occupation as the aggregation of jobs or similar work situations regarding the activities performed, with or without employment relationship (MTE, 2010). It is presented in a hierarchical-pyramidal structure with five levels, ranging from large groups (*'grandes grupos'*) to occupations, as shown in Table 3. Its structure is based on the *International Standard Classification of Occupations* (ISCO). At the finest level, there are 2,570 occupations, also uniquely identified by a six-digit code. Its most recent update was in 2002, the *CBO-2002*, and is the one taken into consideration in this paper.

The large groups structure of the *CBO-2002* follows the structure present in the *ISCO-88*, the *International Standard Classification of Occupations*, that groups occupations based on 'skill levels' and educational attainment. *ISCO-88* uses schooling degree as a measure for skill level, sorting different educational levels into four distinct skill levels, in general terms: (1) *primary education*; (2) *secondary education*; (3) *first stage of tertiary education*; and (4) *second stage of tertiary education*. In this fashion, considering the high skill levels and education attainment STEM occupations usually require[5], it is expected that a solid classification of STEM occupations comprises only the large groups of skill levels 3 and 4, that is, workers with post-secondary education.

[5]In Brazil, it is very common for people who finish high school to enhance their educational attainment with a Certified Course or Program, called 'technical courses'. These 'technical courses' are recognized as post secondary education in the country.

Table 3: CBO-2002 Structure

Structure	# Digits	# Occupations
Grandes Grupos (Large Groups)	1	10
Subgrupos Principais (Main Subgroups)	2	49
Subgrupos (Subgroups)	3	195
Famílias ou Grupos de Base (Families or Base Groups)	4	622
Ocupações (Occupations)	6	2,570

Source: *Ministério do Trabalho Emprego*, 2010.

2.4 Comparison Between Classification Systems

If both classification systems are akin regarding the criteria adopted for their finest level of aggregation, there are noticeable differences between the two classification systems. First, from the much higher count of occupations of the Brazilian classification, it is evident that the degree of detail in the *CBO-2002* exceeds that in *SOC-2010*. Secondly, even though they present similar coding structures, the two systems differ in their leveling criteria, specially for considering the higher levels of aggregation.

As detailed above, while the *CBO-2002* most aggregated level (large groups) sorts occupations by skill level and similarity in the activities performed, the *SOC-2010* does so by work field or domain. There are 10 *CBO-2002* large groups contrasting with the 23 *SOC-2010* major groups from the highest level of aggregation.

3 Matching Between Systems

In order to classify the *CBO-2002* occupations into STEM occupations following the *Attachment C* from USBLS (2012c), it is first necessary to develop a tool to compare occupations between both classification systems. Ideally, we would be able to match each of the occupations listed in *SOC-2010* with a counterpart from the *CBO-2002*. Unfortunately, to our knowledge, there is no available document that fulfills this, so that we must propose a new path. To accomplish this, we made attempts at different methods, listed below.

1. Crosswalk between *SOC-2010* and *CBO-2002*.

2. Six-Digit *CBO-2002* Hand-Coding.

3. **Four-Digit *CBO-2002* Hand-Coding.**

Attempts (1) and (2) were not followed through with. Their detailed classification methods can be found in the Appendices of this paper. The classification method we found to be best is indicated by (3), the Four Digit *CBO-2002* Hand-Coding, for reasons that will become clear in this section.

At attempt (1), we first tried to match *CBO-2002* six-digit occupations to *SOC-2010* six-digit occupations through conversion tables that match the national occupational classification systems of each country to the *International Standard Classification of Occupations* (ISCO) as a crosswalk. This process, however, proved to be inefficient. The different classification systems have different levels of aggregation, so that the conversion between one and the other, at each stage, implied a substantive loss of variation. In addition, this method bears the natural losses of conversion processes in general, such as noise, absence of exact correspondence and errors. For that matter, we have not followed up on this method. To see further detail on this attempt, we provide a crosswalk spreadsheet with all conversion steps, and the detailed explanation of this method can be found in Appendix A.

Hand-Coding

Given that the crosswalk was a weak attempt at conversion between *CBO-2002* and *SOC-2010*, we believe the ideal procedure to minimize conversion losses to be one that directly corresponds *CBO-2002* occupations to *SOC-2010* occupations, matching them one by one. As previously stated, this could be done automatically if there was a conversion table made available by any accountable classification organization. However, the authors are unaware of such a table.

Consequently, in our following attempt we tried a manual conversion between the two systems. Since we only needed to map STEM occupations in *CBO-2002*, our second attempt at conversion was to try to match each of the occupations marked as STEM in *Attachment C* table into *CBO-2002* occupations. It is crucial to emphasize here that this was not an endeavor at a complete conversion between the two occupational classification systems, such proposed in the crosswalk, but a search for correspondence focusing on STEM occupations.

For this method, we tried different *CBO-2002* aggregation levels to find the best correspondence with *SOC-2010* occupations. We began the hand-coding by scanning through each one of the 2,570 six-digit occupations in *CBO-2002*, what we considered to be the natural path to follow since the *USBLS* STEM definition of occupations also applies to *SOC-2010* six-digit codes. This is indicated by attempt (2). By doing so, we sought to maintain consistency in the degree of granularity of the occupational definitions between each country. However, this method proved to be poor in compatibility, for *CBO-2002* had 2,570 six-digit occupations, against *SOC-2010*'s 840 six-digit detailed occupations, proving to be too narrow a classification (see Appendix B for further detail).

Because of the weak compatibility between six-digit *CBO-2002* and *SOC-2010* occupations, we moved up a notch and attempted hand-coding the four-digit *CBO-2002* occupations (families) instead. This was attempt (3), and what we believe to be the best attempt at conversion considered. We found the degree of detail at this aggregation level to be closer to the six-digit *SOC-2010* than the previous one, and the number of families (see Tables 2 and 3) in *CBO-2002* (622) is more similar to the total of detailed occupations in *SOC-2010* (840) than the number of occupations in *CBO-2002*. Moreover, the *CBO-2002* occupations descriptions at the four-digit level were also more akin to the *SOC-2010* descriptions.

USBLS's *Attachment C* table served as guide for categorizing between STEM and non-STEM occupations. For each of the 622 *CBO-2002* four-digit occupational families, a correspondence was sought among the *SOC-2010* set of 97 STEM occupations listed in *Attachment C* table. If such a match was found, the *CBO-2002* occupation was marked as STEM. The remaining occupational families were considered non-STEM.

Since there were relatively few occupational families in the *CBO-2002*, the whole process was double checked by going in the reverse direction, that is, taking each of the STEM occupations in the *USBLS* document and looking for a *CBO-2002* occupational family counterpart. This allowed us to track and identify more easily potential mistakes like false cognates or occupations that could have more than a single correspondence between the two classification systems.

Even though this classification was straight-forward, differences between the occupations in the two classifications systems still remained and some arbitrariness was required for con-

verting. Henceforth, to eliminate uncertainty when doubt arose, each occupation description from both classification systems was checked and compared.

Table 4: STEM Occupations and Jobs in Brazil, Four-digit Hand Coding Classification

	Brazil (2017)		U.S. (2015)	
	STEM	Total	STEM	Total
# of Occupations	65	622	97	840
	(10.5%)	-	(11.5%)	-
# of Jobs (*in thousands*)	924	32,117	8,620	150,573
	(2.9%)	-	(5.7%)	-

Source: Author's own elaboration based on *U.S. Department of Commerce* using *RAIS*, 2017 and *Ministério do Trabalho e Emprego*.

Table 4 shows the number of STEM jobs and occupations in Brazil according to this classification. There were 65 four-digit STEM occupational families in the *CBO-2002*, out of 622. All of the STEM occupational families belong to the *CBO-2002* large groups ('grandes grupos') 1 (managerial occupations), 2 (science and arts professionals) and 3 (mid-level technicians). As explained in Section 2.3, *CBO-2002*'s large groups structure follow *ISCO-88*, that groups occupations based on 'skill levels' and educational attainment. Thus, only occupations of skill levels 3 and 4 were considered STEM[6], that is, 'Non-academic post-secondary education' and 'Academic education'.

In relative terms, the share of STEM occupations in this classification (10.5%) is close to the share of STEM occupations in *SOC-2010* proposed by the *USBLS* (11.5%). When we look at labor market data, these occupations represent 2.9% of the labor market share, or 924 thousand of workers.

Considering all of the three different classification methods attempted and its possibilities and limitations, we consider the four-digit hand coding classification to be the best fit for conversion between *SOC-2010* detailed occupations and *CBO-2002* occupations. The occupational count in the two level of aggregations (detailed occupations and occupational families) gives a clue about the degree of detail of each aggregation level, in spite of the

[6]Managerial occupations do not have a skill level specification due to the fact that managers have different educational levels and, therefore, heterogeneous skill levels.

Table 5: All proposed STEM classifications: number of occupations, digits and workers

Attempt	Classification	Digits	STEM Occupations	Total Occupations	Workers	Section
(1)	Crosswalk	6	142	2,570	-	Appendix A
			(5.5%)			
	Hand Coding	6	250	2,570	866	
			(9.7%)		(2.4%)	
(2)	HC Aggregation	4	78	622	1,386	Appendix B
			(12.5%)		(3.9%)	
	HC Aggregation	4	48	622	441	
			(7.7%)		(1.2%)	
(3)	**Hand-Coding**	**4**	**65**	**622**	**924**	**Section 3**
			(10.5%)		**(2.9%)**	

Notes: The numbers of workers are reported in thousands. The rows indicated by 'HC Aggregation' in Attempt (2) refer to an aggregation attempt of the six-digit Hand-Coding classification and is detailed in the Appendix B.

Source: Author's own elaboration using *RAIS*, 2017 and *Ministério do Trabalho e Emprego*.

codes having a different number of digits. For comparison, Table 5 displays each classification method attempted, with its respective aggregation level considered (represented by the number of digits) and the number of STEM and total occupations of each method.

Besides the reasons presented above for believing this to be the best fit, the share of STEM occupations in each classification system is very close in this classification method. Finally, converting occupations by hand has reduced information losses by concentrating the conversion process in a single step and allowed a direct comparison between the two classification systems. The full table with all of the STEM four-digit occupational families can be consulted in the Appendix (see Table 22).

Appendix A: *CBO-2002* to *SOC-2010* using *ISCO* Crosswalk

Section 3 mentioned the attempt at performing a complete crosswalk between *CBO-2002* and *SOC-2010*. This Appendix further details and describes carefully the classification method. For this classification, we used the *CBO-2002* six-digit occupations. As stated in Section 3.1, there is no document matching the two classification systems, but there are tables that match them to the *International Standard Classification of Occupations* (ISCO). Thus, the crosswalk was done in four steps, enumerated below.

$$\textit{CBO-2002} \Rightarrow \textit{ISCO-88} \Rightarrow \textit{ISCO-08} \Rightarrow \textit{SOC-2010} \Rightarrow \text{STEM } \textit{USBLS}$$

The first step consists in converting *CBO-2002* to *ISCO-88*[7] (International Standard Classification of Occupations) of the International Labor Organization (ILO), made available by a table from the Brazillian *Comissão Nacional de Classificação CONCLA*. There were multiple occupations in the input classification with the same output code, that is, *CBO-2002*'s unique occupations sometimes had the same correspondent in *ISCO-88*. This makes sense, given that the six-digit *CBO-2002* encompasses a total of 2,750 occupations while the *ISCO-88* only 390 occupations, being less aggregated. This conversion was not complete, since only 1,388 of the 2,570 six-digit occupations were present in the conversion table provided by *CONCLA-IBGE*, just a little over half of the total number of occupations (54%). Furthermore, from the 1,388 *CBO-2002* occupations present in the conversion table, 10 of them did not had an *ISCO-88* correspondent. Altogether, 1,378 *CBO-2002* occupations were converted, which in turn corresponded to 165 unique *ISCO-88* occupations.

The second step was to convert *ISCO-88* to *ISCO-08*.[8] The transformation from *ISCO-88* to *ISCO-08* was done using a table from ILO[9]. As this step was more of an update than a conversion between classifications in fact, and the two versions of the classification came from

[7]There was no table with the *CBO-2002* conversion to most recent *ISCO-08* classification. If available, this would have saved us one less step in the classifications crosswalk.

[8]The *USBLS* did not have a table between *SOC-2010* and the most recent international classification, *ISCO-08*. Consequently, an additional conversion step between the two international classifications versions had to be done.

[9]Available at: http://www.ilo.org/public/english/bureau/stat/isco/docs/correspondence08

Table 6: Step 1: CBO-2002 to ISCO-88

	CBO-2002	ISCO-88
Digits	6	4
Total Occupations	2,570	390
Converted Occupations	1,388	390
	(54%)	
Converted	1,378	295

Note: The rows 'Total Occupations' and 'Converted Occupations' report the number of occupations in each classification and the number of occupations present in the conversion tables, respectively.

Source: *Ministério do Trabalho e Emprego*

the same international organization, this step had no losses, and there were exact matches for each of the occupations. In this step there was more than one correspondence for the occupations of *ISCO-88* in relation to *ISCO-08* and vice versa. In all, *ISCO-88* had 390 occupations, while the *ISCO-08* version is a little broader, with a total of 436 occupations.

Considering the previous step, 1,378 *CBO*'s were converted into 295 unique *ISCO-88* occupations, which in turn were converted into 233 unique *ISCO-08* occupations.

Table 7: Step 2: ISCO-88 to ISCO-08

	ISCO-88	ISCO-08
Digits	4	4
Total Occupations	390	436
	(100%)	(100%)
Converted	295	233

Source: *International Labour Organization.*

The third step was to convert the *ISCO-08* occupations into *SOC-2010*. This was possible from a conversion table by *USBLS* [10]. In this step, there was more than one *SOC-2010* correspondence for the same *ISCO-08* occupations, and vice versa. All *ISCO-08* occupations found a *SOC-2010* match in this step, without losses. The end result was that 1,378 of the

.docx

[10] Available at: https://www.bls.gov/soc/soccrosswalks.htm.

CBO-2002 occupations available in the *CONCLA-IBGE* table were converted into a *SOC-2010* occupation. In all, the *CBO-2002* six-digit occupations were transformed into 191 *SOC-2010* occupations, out of a total of 840 *SOC-2010* six-digits occupations.

Table 8: Step 3: ISCO-08 to SOC-2010

	ISCO-08	SOC-2010
# Digits	4	6
Total Occupations	436	840
	(100%)	(100%)
Converted	233	191

Source: *International Labour Organization* and *US Bureau of Labor Statistics*.

The fourth and final step was to identify which of the *CBO-2002* occupations converted into 191 *SOC-2010* occupations were considered STEM. For this step, we turned to the *Attachment C* table from the *USBLS*. Out of the 840 *SOC-2010* occupations, 97 of them were marked as STEM. Among the converted 191 *SOC-2010* occupations, 27 of them were STEM. Therefore, 70 STEM occupations from *Attachment C* did not find a *CBO-2002* match in this method. On the other hand, among the 1,378 converted *CBO-2002* occupations, 142 (10%) were marked as STEM.

Table 9: Step 4: SOC-2010 to STEM

	SOC-2010	STEM (USBLS)
Total Occupations	840	97
	(100%)	(100%)
Converted	191	27

Source: *US Bureau of Labor Statistics*.

Although it was possible to perform the crosswalk, we consider that the losses were significant. In addition to the *CONCLA CBO-2002* to *ISCO-88* conversion table having only half the occupations of the Brazilian classification, which in itself already represents a huge loss of information, the different levels of aggregation of each classification also discarded relevant information. Moving from 1,388 initial occupations to just 191 different occupations

meant a great loss of variation in the data. Table 10 synthesizes the loss of information in each step of the crosswalk.

Table 10: Crosswalk: all steps

	CBO-2002	ISCO-88	ISCO-08	SOC-2010	STEM (USBLS)
# Digits	6	4	4	6	-
Total Occupations	2,570	390	436	840	97
Converted Occupations	1,388	390	436	840	97
	(54%)	(100%)	(100%)	(96%)	(100%)
Converted	1,378	295	233	191	27
	(53%)	(75%)	(53%)	(23%)	(28%)

Note: The rows 'Total Occupations' and 'Table Occupations' report the number of occupations in each classification and the number of occupations present in the conversion tables, respectively.
Source: *Ministério do Trabalho e Emprego*, *International Labor Organization* and *US Bureau of Labor Statistics*.

We believe that the loss of information from the crosswalk method is mainly due to the changes in the various agregation levels, displayed in Table 11. Starting from the very disaggregated six-digit *CBO-2002*, when transformed into *ISCO* occupations, regardless of the international classification year version, several detailed occupations are grouped under the same code.

Table 11: Classification Systems

Classification	Level of Aggregation	Digits	Occupations
CBO-2002	(Detailed) Occupation	6	2,570
ISCO-88	Unit Group	4	390
ISCO-08	Unit Group	4	436
SOC-2010	Detailed Occupation	6	840

Source: *Ministério do Trabalho Emprego*, *International Labour Organization* and (*US Bureau of Labor Statistics)*.

When we take a step forward and convert the occupations from *ISCO* to *SOC-2010*, we move in the opposite direction of aggregation: starting from a more aggregated classification (*ISCO*, four-digits) to a less aggregated classification (*SOC-2010*, six-digit). Consequently, there is more than one *SOC* correspondence for the same *ISCO* occupation. It is not possible to recover with the *SOC* the initial variety of *CBO* occupations converted into the same *ISCO*.

Table 11 displays the various Classification Systems and its respective number of digits and number of distinct occupations at the finest level of aggregation.

A.1 Crosswalk Example

For a better understanding, we illustrate the loss of information in the process with an example. Starting from the *CBO-2002* occupation *Statistics Professor, post secondary*, indicated by the code 2341-15, the conversion table from the MTE gives us an unique *ISCO-88* occupation match, *College, university and higher education teaching professionals*, represented by the code 2310.

Table 12: Step 1 Example: CBO-2002 to ISCO-88

Classification	Occupation	Code
CBO-2002	*Professor de Estatística no Ensino Superior* (Statistics Professor, Post Secondary)	2341-15
ISCO-88	College, university and higher education teaching professionals	2310

Source: *Ministério do Trabalho Emprego*, International Labour Organization

However, this was not the only occupation converted into the 3131 *ISCO-88* code, due to the less aggregated character of the *ISCO* classifications. In fact, this occupation groups all of higher education teaching professionals in a single occupation, regardless of domain. Another thirty (30) *CBO-2002* occupations were converted into the same *ISCO-88* code, 2310. Some are displayed in Table 13, for illustrating purposes.

Table 13 clearly shows that at this stage there is a huge loss of variation due to the aggregation of data, combining professionals of very different fields into a single occupation. In the next step, there is not much information loss, since there are minor differences between the two *ISCO* classifications. In this particular example, the conversion only does a slight change in the occupation title and the occupation code is the same[11]. Table 14 shows this conversion step below.

The third step depicts how the change in aggregation level affects the conversion efficiency by reducing variation, which in turn results in information loss. Although there is and exact correspondence between the *SOC-2010* occupation *Mathematical Science Teachers,*

[11]This is not true for all *ISCO-88* to *ISCO-08* conversions, though.

Table 13: Step 1 Example: Possible matches between CBO-2002 and ISCO-88

CBO-2002 Code	CBO-2002 Occupation	ISCO-88 Code
2341-15	**Professor de estatística no ensino superior** (Statistics Professor)	2310
2342-05	*Professor de física ensino superior* (Physics Professor)	2310
2342-10	*Professor de química ensino superior* (Chemistry Professor)	2310
2346-48	*Professor de literatura francesa* (French Literature Professor)	2310
2347-05	*Professor de antropologia do ensino superior* (Anthropology Professor)	2310
2347-35	*Professor de filosofia do ensino superior* (Philosophy Professor)	2310
2347-45	*Professor de história do ensino superior* (History Professor)	2310
2347-60	*Professor de psicologia do ensino superior* (Psychology Professor)	2310

Source: *Ministério do Trabalho Emprego*

Note: All Professors in this table are Postsecondary Teachers

Table 14: Step 2 Example: ISCO-88 to ISCO-08

Code	Occupation Title	ISCO Version
2310	College, university and higher education teaching professionals	ISCO-88
2310	University and higher education teachers	ISCO-08

Source: *International Labor Organization*

Post secondary, which would be the best match for our initial *CBO-2002 Statistics Professor* occupation, there are many other occupations with the same *ISCO-08* correspondence code as *Mathematical Science Teachers, Post secondary*. To be precise, there are other 36 *SOC-2010* occupations under the same 2310 *ISCO-08* code. Table 15 illustrates the issue, by presenting some examples.

Table 15: Step 3 Example: Possible matches between ISCO-08 and SOC-2010

ISCO-08 Code	SOC-2010 Occupation	SOC-2010 Code
2310	**Mathematical Science Teachers, Postsecondary**	25-1022
2310	*Business Teachers, Postsecondary*	25-1011
2310	*Computer Science Teachers, Postsecondary*	25-1021
2310	*Architecture Teachers, Postsecondary*	25-1031
2310	*Sociology Teachers, Postsecondary*	25-1067
2310	*Social Sciences Teachers, Postsecondary, All Other*	25-1069
2310	*Health Specialties Teachers, Postsecondary*	25-1071
2310	*Nursing Instructors and Teachers, Postsecondary*	25-1072

Source: US *Bureau* of Labor Statistics.

The problem here is that the various *CBO-2002* occupations that were grouped into a single *ISCO* code, are now requiring a more refined level of aggregation for a good conversion. Since there is more than one possible correspondence to the same *ISCO* code, the automated correspondence process when there are multiple alternatives is not efficient. As a result, the occupation *University and higher education teachers* of the *ISCO-08* is converted into *SOC-2010*'s occupation *Business Teachers, Post secondary*, which points out a distortion of the initial information.

Table 16: Step 3 Example: ISCO-08 to SOC-2010

Classification	Occupation	Code
ISCO-08	University and higher education teachers	2310
SOC-2010	Mathematical Science Teachers, Postsecondary	25-1022

Source: *US Bureau of Labor Statistics.*

However, if we take a look into all of the 37 *SOC-2010* correspondents for the 2310 *ISCO-08* code, we can find the occupation called *Mathematical Science Teachers, Post secondary*,

with the *SOC* code 25-1022, which seems to be a good match for the *CBO-2002* initial *Statistics Professor, Post secondary* occupation. This illustrates how the crosswalk process may lead to inaccuracy and wrong correspondences.

Albeit the occupation *Post secondary Mathematical Science Teachers* seems to be a good match for *Post secondary Statistics Professors* at first glance, if we screen through all of the 840 *SOC-2010* occupations, one can find an occupation called *Staticians*, identified by the code 15-2041. To be really certain of the correct correspondence, the best practice here would be to look at each of the potential matches' descriptions, and search for the best correspondence. Table 17 exhibits the original *CBO-2002* occupation and its possible matches descriptions.

Table 17: Occupations Descriptions and Codes

Classification	Code	Occupation Title	Occupation Description
CBO-2002	2341-15	Statistics Professor, Postsecondary	They teach mathematics, statistics and computing, conduct research, produce academic work in their area of competence; they guide students, plan and implement courses and disciplines, assess student performance, programs and institutions. Coordinate academic and scientific activities. They can provide advice in the technical and scientific areas and collaborate in institutional activities.
SOC-2010	25-1022	Mathematical Science Teachers, Postsecondary	Teach courses pertaining to mathematical concepts, statistics, and actuarial science and to the application of original and standardized mathematical techniques in solving specific problems and situations. Includes both teachers primarily engaged in teaching and those who do a combination of teaching and research.
SOC-2010	15-204	Staticians	Develop or apply mathematical or statistical theory and methods to collect, organize, interpret, and summarize numerical data to provide usable information. May specialize in fields such as biostatistics, agricultural statistics, business statistics, or economic statistics. Includes mathematical and survey statisticians.

Note: The occupations descriptions presented in this table were extracted from *Ministério do Trabalho e Emprego* Book 1 of CBO's Codes, Titles and Descriptions and *US Bureau of Labor Economics* website.

Taking into account the entire process detailed above and the relevant loss of information in each of the steps described, we consider this conversion method to be inefficient. Although it is based on conversion tables elaborated by national and international organizations, it has to deal with severe loss of variation in each of the steps mainly due to the various aggregation levels that the different classifications have, in addition to bearing the natural losses of conversion processes in general (noise, absence of exact correspondence, errors, etc.),

We believe that the ideal procedure to minimize conversion losses is to directly compare the *CBO-2002* occupations with *SOC-2010* occupations, and match them one by one. As noted before, this could be done mechanically if there was a conversion table elaborated by any accountable classification organization. However, as previously stated, the authors are unaware of such a table. Thus, we made new attempts at converting *CBO-2002* into *SOC-2010* occupations, testing different occupation levels of the *CBO-2002* structure to find the best fit. These conversions are detailed in Section 3 and Appendix B.

Appendix B: Detailed Six-Digit Hand Coding

Once the crosswalk conversion process between the *CBO-2002* and the *SOC-2010* classification systems proved to be inefficient, and there was a significant loss of relevant information, we attempted converting the occupations of the two systems manually. Our second attempt at this was to try to match each of the occupations from *CBO-2002* with the occupations marked as STEM from *Attachment C* of the *USBLS* table.

For this classification, we used the six-digit *CBO-2002* occupations. We considered this to be the natural path to follow, since the *USBLS* STEM definition of occupations also used the *SOC-2010* six-digit codes. By doing so, we sought to maintain consistency in the degree of granularity of the occupational definitions between each country. However, this method proved to be poor in compatibility, as will become clear further.

It is crucial to emphasize here that this was not an attempt at a complete conversion between the occupational classification systems of the two countries such as we did previously, but a search for correspondence with a focus on STEM occupations. In this sense, instead of taking each occupation from *CBO-2002* and finding its corresponding *SOC-2010* occupation, we looked at each of the 2,570 *CBO-2002* six-digit occupations and checked if there was any potential match among the *SOC-2010* set of 97 STEM occupations listed in *Attachment C* table. If a match was found close enough, we marked the *CBO-2002* occupation as STEM. The remaining occupations were marked as non-STEM.

It is expected to find differences between the two classification systems and the types of jobs carried out in both countries. For that reason, this type of classification requires some degree of arbitrariness. To minimize the arbitrariness of the classification, some criteria were adopted. As different jobs require different educational levels, only those that required post secondary education – such as a college degree or a high school degree with additional technical or vocational education – were considered STEM[12], to the detriment of occupations that required lower levels of education.

Technologists were classified as STEM for having post secondary education in general, while for technicians only those who had a direct match in *Attachment C* STEM jobs were

[12] In *CBO-2002* 'skill levels' terms, explained in Section 2.3, this would mean skill levels 3 and 4 only.

considered STEM. In the case of teaching professions, only post secondary teachers that taught in STEM fields were marked as STEM occupations. Finally, if after all these steps the comparability between two occupations was still uncertain, the description of the occupations was consulted and compared.

Table 18 shows the number of STEM occupations relative to the total of occupations in each classification. STEM occupations in the *USBLS* table represent 12% of the total of *SOC* occupations (97 out of 840), and 10% of the total of *CBO* occupations (250 out of 2,570) according to this definition.

Table 18: Six-digit STEM Occupations in SOC-2010 and CBO-2002

	CBO-2002		SOC-2010	
	STEM	Total	STEM	Total
# of Occupations	250	2,570	97	840
	(9.7%)	-	(11.5%)	-

Source: Author's own elaboration based on Attachment C (*US Bureau of Labor Statistics)* and *Ministério do Trabalho Emprego.*

Even though both classification codes had the same number of digits, the total number of occupations in this level of aggregation differed greatly. While the six-digit *CBO-2002* comprehends 2,570 unique occupations, the six-digit *SOC-2010* only has 840 single occupations, leading us to believe that the degree of detail in the six-digit occupations of both systems differ significantly. This raised concerns about the comparability between occupations at the finest level of aggregation of the two classification systems.

In spite of the limitations described above, we looked at some labor market data to better understand the classification's implications. To do so, we used the *Registro Anual de Informações Sociais* (RAIS), an administrative data set which contains information of formal workers, collected by the Brazilian Ministry of Labor (MTE). According to this STEM classification, in 2017 Brazil had about 900 thousand STEM workers, much less than the 8.6 million in the United States in 2015 as reported by the *US Department of Commerce* (2017). In relative terms, this represented 2.7 percent of the formal workers in Brazil, nearly half of the US share (5.7%). These numbers are reported in Table 19.

From all of the classifications proposed in this paper, this was the one we believe to most

Table 19: STEM Jobs in the United States and Brazil, Six-digit Classification

	Brazil (2017)		U.S. (2015)	
	STEM	Total	STEM	Total
# of Jobs (*in thousands*)	866	32,117	8,620	150,573
	(2.7%)	-	(5.7%)	-

Source: Author's own elaboration based on *U.S. Department of Commerce* using *RAIS*, 2017.

similar in method to the one in Bonini et al. (2020). The authors also classify the six-digit *CBO-2002* occupations, but follow instead the *U.S. Department of Commerce, Economics and Statistics Administration* (ESA) criteria. Both *USBLS* and *ESA* are accountable and respected US bodies that produce labor statistics. However, their STEM definitions are quite different: while *ESA* classifies only 50 *SOC-2010* occupations as STEM, *USBLS* proposes a classification of 97 STEM occupations. Unsurprisingly, our numbers differed greatly from the ones presented by Bonini et al. (2020)). While the authors consider 164 out of the 2,570 *CBO-2002* occupations as STEM, we found 250 STEM occupations.

B.1 Four-digit Aggregation

To test whether the aggregation level or number of digits taken into account made a meaningful difference in the STEM classification, we disaggregated *CBO*'s occupations to the next higher level, *CBO*'s 'families', represented by a 4-digit code (there is no intermediate 5-digit CBO). In all, there are 622 families in *CBO* (check Table 3).

As *CBO* families consist of groups of various six-digit occupations, consequently there were families with mixed occupations between STEM and non-STEM. More specifically, there when aggregated into families, the previous six-digit occupations formed three sets of families: (i) families composed of only STEM occupations; (ii) families composed of some STEM occupations and some non-STEM occupations; and (iii) families composed of only non-STEM occupations.

Hence, we tested two new classifications, one in which for the family to be considered STEM *all occupations* within that family should be STEM occupations, and another in which *at least one* occupation within that family needed to be STEM for that family as a

whole to be considered STEM. Namely, sets (i) and (ii) above, respectively.

In the first of these new classifications, presented in Table 20, 78 families were considered STEM, which represents 12% of the total families, a share of STEM occupations very close to that documented by the *USBLS*. Nevertheless, this classification seemed a little noisy, since it considered some families that had only one occupation considered STEM among many others that were not related to the field, therefore not being representative of the work domains of that family. For instance, this classification included families of clinical and veterinary doctors, professions that are not considered STEM in general. In terms of the labor market, this classification categorized almost 1.4 million workers as STEM (4.3% of the labor market) in 2017.

Table 20: STEM Occupations and Jobs in Brazil, Four-digit 'At least one STEM' Aggregation

	Brazil (2017)		U.S. (2015)	
	STEM	Total	STEM	Total
# of Occupations	78	622	97	840
	(12.5%)	-	(11.5%)	-
# of Jobs (*in thousands*)	1,386	32,117	8,620	150,573
	(4.3%)	-	(5.7%)	-

Source: Author's own elaboration based on *U.S. Department of Commerce* using *RAIS*, 2017 and *Ministério do Trabalho e Emprego*.

In the second classification, considering families composed of *only* STEM six-digit occupations, 48 families were classified as STEM. As Table 21 shows, this is equivalent to 8% of the total families. Unlike the previous classification, this in turn seemed to be very restricted, disregarding families that were clearly in the STEM domains, like post secondary teachers from related areas, some families of technicians, drafters, and information technology managers, for example. In the labor market, they accounted for only 440 thousand jobs. Of the total formal jobs, only 1.4% are STEM by this classification.

Considering the two proposals for disaggregation above, the two classifications have opposite problems: while one seems too broad, the other seems too restrictive. However, *CBO*'s four-digit occupations seemed closer to *SOC*'s six-digit occupations, both in terms of the number of occupations (622 to 840) and the similarity between the names of occupations at

Table 21: STEM Occupations and Jobs in Brazil, Four-digit 'All STEM' Aggregation

	Brazil (2017)		U.S. (2015)	
	STEM	Total	STEM	Total
# of Occupations	48	622	97	840
	(7.7%)	-	(11.5%)	-
# of Jobs (*in thousands*)	441	32,117	8,620	150,573
	(1.4%)	-	(5.7%)	-

Source: Author's own elaboration based on *U.S. Department of Commerce* using *RAIS*, 2017 and *Ministério do Trabalho e Emprego*.

this level of aggregation. In other words, the similarity between *CBO*'s family names and *SOC* occupations appears to be greater than the similarity between the detailed occupations of *CBO* and *SOC*. In this sense, we were led to believe that we should continue with a four-digit classification, as it seems the most appropriate level of aggregation. The ideal classification of STEM and non-STEM families in *CBO-2002* should be somewhere between the two classifications tested and described above, neither too broad nor too restricted.

Lastly, we took a step forward and did the same test with the *CBO*'s 3-digit classification. This classification proved to be very aggregated, since the *CBO* has only 195 three-digit subgroups. Only 8 subgroups were composed entirely of detailed occupations (six-digit) of STEM fields, and 29 subgroups with at least one six-digit STEM occupation. It can be said that this aggregation only amplified the problems of aggregation from the previous level, since it further restricted the sets entirely composed of STEM occupations in the strictest classification, and included even more non-STEM six-digit occupations in the broader aggregation. Thus, we did not look at labor market data for these two classifications.

Appendix C: *CBO-2002* STEM Occupational Families

Table 22: STEM CBO-2002 Occupational Families

Occupational Family	Code
Diretores de operações de obras em empresa de construção	1223
Diretores de tecnologia da informação	1236
Diretores de pesquisa e desenvolvimento	1237
Gerentes de obras em empresa de construção	1413
Gerentes de tecnologia da informação	1425
Gerentes de pesquisa e desenvolvimento e afins	1426
Profissionais da biotecnologia	2011
Profissionais da metrologia	2012
Engenheiros de controle e automação, engenheiros mecatrônicos e afins	2021
Pesquisadores das ciências biológicas	2030
Pesquisadores das ciências naturais e exatas	2031
Pesquisadores de engenharia e tecnologia	2032
Pesquisadores das ciências da saúde	2033
Pesquisadores das ciências da agricultura	2034
Profissionais da matemática	2111
Profissionais de estatística	2112
Engenheiros em computação	2122
Administradores de tecnologia da informação	2123
Analistas de tecnologia da informação	2124
Físicos	2131
Químicos	2132
Profissionais das ciências atmosféricas e espaciais e de astronomia	2133
Geólogos, oceanógrafos, geofísicos e afins	2134
Engenheiros ambientais e afins	2140

Continued on next page

Table 22 – *Continued from previous page*

Occupational Family	Code
Engenheiros civis e afins	2142
Engenheiros eletricistas, eletrônicos e afins	2143
Engenheiros mecânicos e afins	2144
Engenheiros químicos e afins	2145
Engenheiros metalurgistas, de materiais e afins	2146
Engenheiros de minas e afins	2147
Engenheiros agrimensores e engenheiros cartógrafos	2148
Engenheiros de produção, qualidade, segurança e afins	2149
Biólogos e afins	2211
Biomédicos	2212
Engenheiros agrossilvipecuários	2221
Engenheiros de alimentos e afins	2222
Professores de matemática, estatística e informática do ensino superior	2341
Professores de ciências físicas, químicas e afins do ensino superior	2342
Professores de arquitetura e urbanismo, engenharia, geofísica e geologia do ensino superior	2343
Professores de ciências biológicas e da saúde do ensino superior	2344
Profissionais em pesquisa e análise geográfica	2513
Técnicos em mecatrônica	3001
Técnicos em eletromecânica	3003
Técnicos de laboratório industrial	3011
Técnicos de apoio à bioengenharia	3012
Técnicos químicos	3111
Técnicos de produção de indústrias químicas, petroquímicas, refino de petróleo, gás e afins	3112
Técnicos em controle ambiental, utilidades e tratamento de efluentes	3115
Técnicos em construção civil (edificações)	3121

Continued on next page

Table 22 – *Continued from previous page*

Occupational Family	Code
Técnicos em construção civil (obras de infraestrutura)	3122
Técnicos em geomática	3123
Técnicos em geologia	3161
Técnicos de desenvolvimento de sistemas e aplicações	3171
Técnicos de suporte e monitoração ao usuário de tecnologia da informação.	3172
Desenhistas técnicos, em geral	3180
Desenhistas técnicos da construção civil e arquitetura	3181
Desenhistas técnicos da mecânica	3182
Desenhistas técnicos em eletricidade, eletrônica, eletromecânica, calefação, ventilação e refrigeração	3183
Desenhistas projetistas de construção civil e arquitetura	3185
Desenhistas projetistas da mecânica	3186
Desenhistas projetistas da eletrônica	3187
Técnicos em biologia	3201
Técnicos florestais	3212
Técnicos de apoio à biotecnologia	3253
Técnicos de apoio em pesquisa e desenvolvimento	3951

References

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in stem: A gender gap to innovation. *Economics and Statistics Administration Issue Brief* (04-11).

Bonini, P., Custódio, C. F., da Silva, F., & Okawati, G. A. A. (2020). Formação e atuação profissional nas áreas stem no brasil: Ainda temos pouco? In W. L. Linhares (Ed.), *Desafios da educação técnico-científica no ensino médio* (Vol. 2, pp. 252–263). Atena Editora.

ILO. (2012). *International standard classification of occupations 2008 (isco-08): structure, group definitions and correspondence tables*. International Labour Office.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). Stem: Good jobs now and for the future. esa issue brief# 03-11. *US Department of Commerce*.

Lütkebohle, I. (2008). *BWorld Robot Control Software.* ([Online; accessed 19-July-2008])

MTE. (2010). *Classificação brasileira de ocupações. livro 1. códigos, títulos e descrições* (3rd ed.). https://wp.ufpel.edu.br/observatoriosocial/files/2014/09/CBO-Livro-1.pdf.

Muendler, M.-A., Poole, J., Ramey, G., & Wajnberg, T. (2004). Job concordances for brazil: Mapping the classificaçao brasileira de ocupaçoes (cbo) to the international standard classification of occupations (isco-88). *University of California, San Diego, unpublished manuscript*.

Noonan, R. (2017). Women in stem: 2017 update. esa issue brief# 06-17. *US Department of Commerce*.

Schwartzman, S. (2018). Educação técnico científica no brasil: Desafios e perspectivas. In D. Foguel & M. Scheuenstuhl (Eds.), *Desafios da educação técnico-científica no ensino médio* (Vol. 1, pp. 11–20). Academia Brasileira de Letras.

Seemann, M. W., & Bonini, P. (2017). Trabalho stem no brasil de acordo com a cbo. In *27º seminário de iniciação científica.* UDESC.

USBLS. (2010). *2010 soc user guide.* https://www.bls.gov/soc/soc_2010_user_guide.pdf.

USBLS. (2012a). *Attachment a: Soc policy committee recommendation to omb. options for defining stem (science, technology, engineering, and mathematics) occupations under the 2010 standard occupational classification (soc) system.* https://www.bls.gov/soc/Attachment_A_STEM.pdf.

USBLS. (2012b). *Attachment b: Stem definition options.* https://www.bls.gov/soc/Attachment_B_STEM.pdf.

USBLS. (2012c). *Attachment c: Detailed 2010 soc occupations included in stem. options for defining stem (science, technology, engineering, and mathematics) occupations under the 2010 standard occupational classification (soc) system.* https://www.bls.gov/soc/Attachment_C_STEM.xls.

USBLS. (2018). *2018 soc user guide. what's new in the 2018 soc.* https://www.bls.gov/soc/2018/soc_2018_whats_new.pdf.