

# THE OCCUPATIONAL CONTEXT OF MISMATCH: MEASURING THE LINK BETWEEN OCCUPATIONS AND AREAS OF STUDY TO BETTER UNDERSTAND JOB- WORKER MATCH<sup>1</sup>

ALEXANDRA MARIN

SEAN HAYES

**Abstract.** Scholars have long been interested in the prevalence, causes, and consequences of workers being well matched or poorly matched to their jobs. Researchers have moved beyond thinking of mismatch as a simple issue of deficit or surplus of skill or education to ask if workers' skills or education are relevant to their jobs. The next step for workers studying job-worker match is to consider the relevance of relevance. The causes and consequences of not having relevant education will be different in occupations that are closely tied to particular fields of study than in those not linked to any field of study. To facilitate this research agenda we develop seven measures of the link between occupations and fields of study in the Canadian labour market. We test the validity and robustness of these measures. We discuss when each measure is most appropriate and provide an appendix listing values for the three best-performing measures, calculated for Statistics Canada 4-digit occupational codes.

**Keywords:** Occupations, Job-Worker Mismatch, Occupational Codes, Fields of Study

*Résumé.* Les sociologues s'intéressent depuis longtemps à la prévalence, aux causes et aux conséquences de l'appariement favorable ou non de travailleurs

1. This research was funded by the Social Sciences and Humanities Council of Canada Grant and conducted using Statistics Canada data in the Statistics Canada Toronto Research Data Centre. Remaining acknowledgments redacted. We are grateful for assistance advice and feedback from Stephen DiPede, Roberto Fernandez, Sherri Klassen, and Scott Milligan. This research was previously presented at the Annual Meeting of the American Sociological Association and the MIT-Harvard Economic Sociology Workshop. We thank attendees for their comments.

à leurs emplois. Ce décalage n'est aujourd'hui plus perçu comme un simple problème de lacune ou de surplus de compétences ou d'éducation, mais plutôt comme une question de pertinence des compétences ou de l'éducation à l'emploi demandé. Il convient donc d'examiner la pertinence même de la pertinence. Les causes et les conséquences d'une éducation inadéquate seront différentes pour les professions nécessitant des études précises par rapport aux professions sans champ d'études spécifique. Pour approfondir ce programme de recherche, nous développons sept mesures du lien entre les professions et les champs d'études dans le marché du travail canadien. Nous discutons des situations où une mesure est jugée plus appropriée qu'une autre, et nous fournissons une annexe énumérant les valeurs pour les trois meilleures mesures, calculées en fonction des codes des professions de Statistiques Canada.

**Mots clés:** Professions, Incompatibilité travail-travailleur, Codes professionnels, Domaines d'études

## INTRODUCTION

Researchers have long been interested in studying the link between the level or type of education and the jobs held by graduates, focusing especially on the extent to which workers are well-matched or mismatched to their jobs (Allen and van der Velden 2001, Handel 2003, Hersch 1991, Kalleberg 2007, Layne 2010). Implicitly, such research suggests that a mismatch – when employees have more or less education than required for the job, or when the employees were trained in a field of study not directly aligned with their position – demonstrates a failing either in the labour market or in the education system. The issue of match is, however, more complicated than a simple examination of correspondance between field of study and field of employment can show. In particular, the significance and relevance of this type of match vary across occupations.

The measures proposed here tease out the significant differences between occupations that can make understanding the match between field of study and occupation upon graduation meaningful. The key distinction between occupations is the level to which the occupation is associated with specific types of education. Some occupations employ only workers with particular narrowly-defined types of education. In such occupations, being mismatched – defined as the absence of this education – is a meaningful concept likely to have significant consequences. Other occupations draw from a broad pool of workers with many types of educational backgrounds. In these occupations no particular background may constitute a strong match, making the concept of mismatch less

meaningful within these occupations. We make this distinction possible by developing measures that indicate where specific occupations in the Canadian labour market fall on this continuum and testing the validity and reliability of these measures.

#### **ALL MISMATCH IS NOT CREATED EQUAL**

While job-worker match can be defined to include a broad range of factors, such as preferences and needs related to compensation, work hours, or location (Kalleberg 2007), the literature most commonly defines match and mismatch in terms of education or skill. Mismatches are divided into cases of overqualification/overeducation, which exist wherever workers' education is not used or required in their work, and underqualification/undereducation, which exists where workers do not have the education or skills required to do their jobs (Allen and van der Velden 2001, Handel 2003, Kalleberg 2007, Sloane 2002). Researchers studying mismatches based on workers' skills rather than education have similarly measured skill as existing in excess or deficit (Allen and van der Velden 2001, Handel 2003, van der Werfhorst 2002).

These understandings of job-worker match typically focus on the *level* of education or skill that jobs require and that workers in those jobs have, defining workers as mis-matched when these are not equal or similar (Hersh 1991, Ortiz and Kucel 2008). However, while level of education is important, many jobs require not only particular levels of education, but also more specialized education (Bills 2003). A more nuanced understanding of job-worker match requires thinking not only about levels of education or skill but about types of education and skill (Robst 2007, 2008). It is possible for a person to both be overeducated for a job and also underqualified. This would be the case if the authors of this paper were to be hired as electricians, an occupation that requires lower levels of education than we possess but also requires specific skills and credentials that we do not have.

Treating all occupations as equivalent fails to problematize the question of how easy it is for a worker's skills or credentials to be well-matched or poorly matched to a particular occupation. Occupations with narrowly-defined educational or training requirements make it possible for persons who have the required credentials to be well-matched to those jobs and for people without them to be mismatched. However, many occupations are not closely tied to any particular field of study. In these occupations it is not possible for one's credentials to be well-matched to the occupation and it does not make sense to treat workers

whose credentials are not relevant to their work as mismatched. To fully appreciate the context of job-worker matches, analysts need to recognize this distinction between occupations. Yet in the absence of a means of separating one type of occupation from the other, researchers have ignored this distinction (e.g. Allen and van der Velden 2001, Robst 2007, Bouderbatt and Chernoff 2009), even considering workers mismatched to their jobs if they state that a field of study other than their own is the most appropriate preparation for their work *or if they state that no particular field of study is relevant* (Allen and van der Velden 2001). They do this in spite of research that has found that graduates of “field-specific” programs, those tightly linked to an occupation or industry, are more likely to be working in jobs related to their education, suggesting both that match between education and occupation is a strong possibility in these fields, and that graduates of non-field-specific programs are not necessarily mismatched in occupations not related to their fields of study (Robst 2007; Heijke, Meng and Ris 2003 van der Werfhost 2002).

#### MEASURING THE LINK BETWEEN FIELDS OF STUDY AND OCCUPATIONS

This article proposes and tests three measures that can be used to assess the extent to which occupations are linked to particular fields of study. Creating such measures requires data about people in a wide variety of occupations and their full educational histories, including fields of study. In Canada these data are available through the National Graduates Surveys (Statistics Canada 1998). However, access to field-of-study-related variables is restricted. Therefore, it would be impractical for individual researchers to each recalculate measures required for individual research projects. We include in an appendix measures for occupations using Statistics Canada standard occupational codes, allowing researchers without access to these restricted data to analyze this link when working with any data set that includes standard occupational codes.

By studying the diversity of fields of study pursued by actual job incumbents, our measures assess the link between occupations and fields of study as it is negotiated by job seekers and employers. We measure the diversity rather than the legal requirements or the extent to which the duties of an occupation are related to the content taught in different educational fields because the credentials or licenses that qualify people to work in particular occupations are not inherent to the occupations and their duties. Expected qualifications for particular occupations are socially constructed and negotiated over time (Murphy 1984). Occupations that were once practiced freely or even illicitly are now regulated profes-

sions (Bourgeault 2000); occupations once requiring no credentials now require university degrees (Collins 1979, van der Werfhorst and Anderson 2005), and occupations once requiring no particular field of study now draw some of their occupants from post-secondary programs dedicated to preparing students for those occupations (Chillas 2010). If the likelihood of securing jobs in an occupation varies with field of study, this matters, regardless of whether these requirements are formally stated by employers or the education content is necessary for performing job duties. This method has the further advantage of neither requiring the time and expertise of trained job analysts, nor permitting the subjectivity and non-reliability of worker self-reports (Ortiz and Kucel 2008). Therefore, examining this link requires not the study of job duties and or the knowledge required for those duties – what organizational researchers call *task knowledge requirements* (Carley 2006) – though these will surely be factors in creating the link, but rather the actual employment patterns that show who works in which occupations and the fields of study from which these workers are drawn.

In the remainder of this article we introduce our data, describe the calculation of our measures and then examine the face validity and predictive validity of those measures. Because calculations of occupation-field-of-study link are based on more survey respondents for some occupations than for others, we test the robustness of our measures to calculation using small numbers of respondents. Findings show that some measures are predominantly sensitive to the dominance of the single most common field of study in an occupation, while others also vary with the distribution of workers among less common fields of study. After identifying the best performing measures in each of these two classes, we discuss the conditions under which each might be useful and discuss the contribution made by the development of these measures.

## DATA AND METHODS

Data for this paper were taken from Statistics Canada's National Graduate Survey (1995 cohort), conducted in 1997 and the follow-up survey conducted in 2000<sup>2</sup>. The surveys investigate the labour market and education experiences of graduates from universities, community colleges, and trade or vocational programs in Canada. Recent graduates are the most appropriate workers for studying the relationships between fields

---

2. We use the 1995 cohort because these data contained many more Job-FOS pairs resulting in a data set with more respondents per occupation, which we show is important in ensuring the reliability of these measures.

of study (hereafter FOS) and work because this is the point in workers' careers when field of study is most likely to influence occupations entered. If an occupation requires no particular background for entry-level workers, it is unlikely to require a particular field of study as those workers gain experience.

The study population consists of graduates of Canadian post-secondary educational institutions who completed the requirements for a degree, diploma or certificate in 1995. This includes graduates of university programs that award certificates, diplomas or bachelors, masters or doctorate degrees. It also includes graduates of community colleges, technical schools and skilled trades programs lasting three months or more. Graduates from private post-secondary institutions, part-time trade courses, apprenticeship programs or vocational programs lasting less than three months were not included, nor were students who completed continuing education courses not leading to a diploma, certificate or degree. The survey employed a stratified, systematic, random sample design. The sample was stratified first by province and then within each province by five levels of education and nine areas of study for the university and technical programs and eight areas of study for the trade and vocational programs. The sample size for each stratum was set to allow for useful levels of detail and equal reliability for every province, level of education and field of study. Computer assisted telephone interviews were conducted with 68.7% of the 61,759 graduates selected for inclusion in the sample.

The survey contains data on every credential attained and every job held since the start of the focal education program. We limit our analyses to the 63,165 jobs that respondents held between 1990 and 2000. Each job is coded using Statistics Canada's 1991 standard four-digit National Occupational Code. For each credential held by respondents, fields of study (hereafter FOS) are coded using USIS and CCSIS (Statistics Canada 1998) major field of study codes for universities and colleges or trade schools respectively. Up to two fields of study were recorded for the 1995 graduation for each post-secondary program completed prior to 1995 and between 1995 and 2000, when the second follow-up survey was conducted. College and university major codes were harmonized and combined by Statistics Canada into a single set of codes consisting of 102 possible fields of study. Respondents completed degrees with a total of 48,709 fields of study.

For each respondent we compiled a list of every possible pairing of a job they had held and a FOS in which they had completed a credential. For example, A respondent who had held one job as an accountant and one job as a receptionist and graduated with a double major in English

and business, would have four job-FOS pairs listed: accountant-English, accountant-business, receptionist-English and receptionist-business. Using the hiring dates and graduation dates for each job and each credential, we eliminated all job-FOS pairs in which a person was hired before graduating with the FOS listed. Using this list we determined the ranking of every FOS represented within each occupation. It was frequently the case that respondents had more than one FOS completed before being hired in a job, either because they had multiple credentials or because they had pursued more than one field of study in completing their degrees. In these cases we considered for each respondent only the FOS that was most common among workers in the relevant occupation<sup>3</sup>. In the example cited above, if business majors are more common among accountants than English majors, and English majors are more common among receptionist than business majors, we would retain the 'accountant-business' listing and the 'receptionist-English' listing and eliminate the others. This resulted in 58,243 job-FOS pairs.

We imported this list into *ORA*, software for social network analysis, weighting each row of data using Statistics Canada assigned respondent weights. The result was a two-mode network, with occupations and FOS as nodes and the weight of connections between these nodes corresponding to the sum of respondent weights for respondents who had studied that FOS and worked in that occupation. For example, if two respondents hold medical degrees and jobs as bank tellers and those two respondents had assigned weights of 1.5 and 3.5, the link between bank tellers and medical degrees would be valued at 5. We can interpret the values attached to each occupation-FOS dyad as being the number of people working in that occupation with that FOS in a representative sample where  $N$  is the sum of respondent weights for respondents included in the data.

We removed from these data those occupations that included fewer than five respondents within the survey sample. This limit is used both to

---

3. Initially we calculated measures using all FOSs rather than one FOS per job holder we report measures calculated with a single FOS per job for a number of reasons. First, the measures were more difficult to interpret intuitively when the numbers on which they were based were worker-FOS pairs rather than workers. Second, the face validity of measures calculated in this way was low particularly around professions requiring graduate study, which appeared to be quite weakly linked to FOSs simply because pre-professional undergraduate degrees can be in a wide variety of fields. This made occupations like "lawyer" appear to be more weakly linked to FOS than professional occupations like pharmacy pursued at the undergraduate level. Finally, results of further analyses showed that measures using only the most common FOS were more robust to small sample sizes and had higher predictive validity.

eliminate occupations for which results were likely to be unreliable and to ensure compliance with Statistics Canada data reporting rules. Eliminating occupations with fewer than five respondents resulted in a data set representing 476 occupations and 102 FOSs. Exporting these data to standard statistical software gave us a data set in which occupations were the units of analyses and each FOS was a variable. Values for each FOS variable indicated the number of people one would expect to observe with that occupation and FOS given a fully representative sample.

### *Calculating the Measures*

We developed three types of measures, each designed to approach the link between occupations and areas of study differently. The first set of measures is based on the proportions of workers with backgrounds in the most popular FOSs. We take the weighted proportion of workers in the most common as well as the 3, 5, and 10 most common FOSs within that occupation. This measure can vary between 0 and 1, with higher values indicating more workers in the most common FOS(s) and thus a stronger link between occupation and education.

The next set of measures includes two measures based on all FOSs represented within an occupation, rather than just the most common. These measure how evenly workers are distributed across FOSs, with more even distributions reflecting weakly linked occupations and greater concentration within some FOSs indicating a strong occupation/education link. The *index of diversity (ID)* is the probability that two randomly selected workers in an occupation do not share their most common FOSs (Lieberson 1969). Our ID-based measure used here is created by subtracting the ID from 1 to give the probability that a randomly selected pair of workers *do* share their most common field of study. The value of this measure increases as the number of FOSs represented within the occupation increases and as the evenness of distribution across FOSs increases. This measure also varies between 0 and 1, with higher numbers indicating greater concentration in a smaller number of FOSs and thus a stronger link. The *Index of Qualitative Variation (IQV)* is a transformation of the index of diversity. It is the index of diversity normalized by division by the highest possible value that the index of diversity could take *given the number of FOSs represented within the occupation*. Because it is normalized in this way, the index of qualitative variation increases when person-FOSs are more equally distributed across the FOSs represented, but not when more FOSs are represented within the occupation. Not accounting for the number of FOSs represented in an occupation discounts an important aspect of occupations' link to fields of study,



but makes this measure less sensitive to variations caused by occupations represented by few respondents for which the number of fields of study that can be represented within the data is naturally limited. Larger values for this measure indicate a strong link between occupations and fields of study.

The third type of measure is network-based, treating people and FOSs as nodes and examining the structure of the ties between them. To calculate *Degree Centralization* we create one network for each occupation. Each network includes people and FOSs as nodes that are connected by a tie weighted with the respondent weight if that person has that FOS as their most common-within-occupation FOS. For example, a respondent who had attended university double majoring in English and Psychology and who was weighted 3.2 within the survey results in the FOS psychology being linked to 3.2 persons<sup>4</sup>. For each network we calculate the extent to which the network is dominated by a single node — that is the extent to which one FOS is more common than *each* other FOS (Freeman 1979). Because there are two kinds of nodes in these data (people and FOS), we use single-mode degree centralization, a measure designed to take this data structure into account when normalizing the measure (Borgotti and Everett 1997, Everett and Borgotti 2005). We use the formula:

$$\frac{\sum_{i=1}^{n_1} (c_{1*} - c_{1i})}{(n_1 - 1)(n_0 - 1)}$$

where  $c_{1*}$  is the number of workers in the occupation who have studied the most common FOS,  $c_{1i}$  is the number of workers in the occupation who have studied in each other FOS,  $n_1$  is the number of FOSs represented in the occupation and  $n_0$  is the weighted number of workers in the occupation. This measure looks at how much more common the most common FOS is than each other FOS and divides this sum by the maximum value that it could take, given the number workers in the occupation and the number of FOSs represented within the occupation. This measure should be closely related to the proportion of workers in the most common FOS, but is better able to distinguish between occupations in which there is a single domin-

4. Strictly speaking the number of nodes in a network is always a natural number; however, weighting the data using the correct survey weights requires us to think about one person representing non-integer numbers of people. This inconsistency could be resolved by multiplying the number of nodes by 100, however, this would not produce substantively different findings since all values would be multiplied by the same constant.

ant field of study and those in which there are a small number of dominant FOSs. Its value is varies based on both the proportion of workers in the most common FOS *and* the distribution of workers across the remaining FOS. Larger values for this measure indicate a stronger occupation/education link.

The measures calculated here are ratio-level measures, for which it is possible to meaningfully and consistently interpret intervals of equal size to indicate equal differences in the occupation-FOS link. Because these are ratio rather than ordinal measures they can be treated as continuous when estimating regression models, and the magnitude of resulting coefficients – which would be meaningless if estimated with ordinal measures – can be meaningfully interpreted and compared (Sørensen 1979).

FINDINGS

Table 1: Bivariate Correlations Between Measures

	PropTop1	PropTop3	PropTop5	PropTop10	ID	IQV	Degree	Centralization
PropTop1	--							
PropTop3	0.845**	--						
PropTop5	0.731**	0.957**	--					
PropTop10	0.566**	0.808**	0.915**	--				
ID	0.967**	0.847**	0.738**	0.569**	--			
IQV	0.906**	0.688**	0.572**	0.413**	0.923**	--		
Degree	0.953**	0.728**	0.603**	0.449**	0.895**	0.889**	--	
Centralization								--

\* p<.05 \*\* p<.01 \*\*\*p>.001

We calculated the link between occupations and FOSs of 476 occupations using each measure. Table 1 shows correlations between measures. Correlations are substantively large, statistically significant and in the expected directions suggesting that the measures are capturing the same underlying phenomenon. The especially strong correlations between ID, IQV, degree centralization, and the proportion of workers in the most single most common FOS, suggest that all of these measures are especially attuned to the same aspect of occupation/education link, the prominence of the most common FOS. Looking at correlations within measures of the proportion of workers with the most one, three, five and ten common FOSs reveals a distinction between the measure based on the single most

common FOS and those based on the three, five or ten most common FOSs. While those based on multiple FOSs are strongly correlated to one another, their correlations with the proportion of workers in the most common FOS and with the other measures capturing the popularity of the most common FOS are weaker. Correlations between measures that vary strongly with the most common FOS and measures based on the popularity of multiple fields of study get weaker as the number of fields of study considered increases. This is expected given that the most popular field of study is an increasingly small component of each measure, but the steepness of the decline and the lack of a similarly steep decline in correlations between the measures looking at the 3, 5 and 10 most common FOSs, suggests again that the measures can be divided into two subsets, each attuned to different aspects of the link between FOS and occupation: ID, IQV, degree centralization and the proportion in the single most common FOS vary primarily with the popularity of the most common FOS, while the proportion of workers in the 3, 5 and 10 most common FOSs are attuned to the broader distribution of FOSs within the occupation.

**Table 2: Descriptive Statistics for Measures of Occupational Closure and Occupational Closure**  
**Values for Selected Occupations, NGS 1995 Cohort. Common**

NOC	Occupational Title	PropTop1	PropTop3	PropTop5	PropTop10	ID	IQV	Degree Centralization
Graduate Professions								
3112	General Practitioners and Family	0.817	0.905	0.943	0.989	0.673	0.651	0.805
4112	Lawyers and Quebec Notaries	0.868	0.937	0.968	0.982	0.756	0.747	0.863
3114	Veterinarians	0.918	0.979	0.996	1.000	0.846	0.815	0.903
Direct Entry Professions								
3152	Registered Nurses	0.867	0.937	0.961	0.980	0.755	0.746	0.862
3222	Dental Hygienists and Dental	0.673	0.991	1.000	1.000	0.527	0.409	0.591
Trades								
6271	Hairstylists and Barbers	0.877	0.935	0.980	1.000	0.772	0.746	0.864
7241	Electricians (Except Ind'l and	0.878	0.946	0.974	0.990	0.775	0.760	0.871
Optionally Credentialed								
6241	Chefs	0.755	0.865	0.905	0.979	0.578	0.550	0.739
112	Human Resources Managers	0.687	0.860	0.927	0.984	0.491	0.459	0.669
6431	Travel Counselors	0.646	0.855	0.925	0.968	0.443	0.412	0.627
Certification-Based								
1113	Securities Agents, Investment	0.353	0.645	0.756	0.915	0.180	0.150	0.326
6231	Insurance Agents and Brokers	0.340	0.529	0.648	0.836	0.152	0.124	0.319
Open Occupations								
1454	Survey Interviewers and	0.193	0.426	0.528	0.693	0.083	0.064	0.177
6421	Retail Salespersons and sales	0.186	0.287	0.370	0.536	0.056	0.043	0.175
6452	Bartenders	0.126	0.277	0.385	0.608	0.051	0.029	0.106
	Mean (all occupations)	0.428	0.691	0.807	0.923	0.720	0.783	0.377
	Standard Deviation	0.197	0.180	0.153	0.098	0.182	0.188	0.202
	N	476	476	476	476	476	476	476

### *Face Validity*

Table 2 shows the mean and standard deviation of each measure, as well as values of the measure for 15 occupations. These occupations were selected for their different relationships to fields of study. We use these occupations to examine the face validity of the measures we have calculated. A measure that is valid on its face would rank occupations requiring particular FOSs as more strongly linked, followed by occupations with related but not required FOSs. Occupations for which related fields of study do not exist should be ranked most weakly linked. We divide occupations into six categories based on their relationship to educational credentials and fields of study.

1. *Graduate Professions* are regulated profession requiring particular programs of study normally undertaken *after* undergraduate study<sup>5</sup>. These occupations vary in the extent to which the first undergraduate programs undertaken are likely to be in varied or uniform areas. For example, physicians are likely to have studied biology, life sciences or other natural sciences in their undergraduate programs, while lawyers may have studied any field but may be concentrated in the humanities and social sciences.

2. *Direct Entry Professions* are also regulated professions requiring particular educational programs for licensing. Professional schools training students for these occupations admit students with only high school diplomas and no previous post-secondary study required.

3. *Trades* are working class occupations that typically require occupation-specific schooling for certification. Schooling required in these occupations is more likely to occur in trade school or apprenticeship programs, not universities.

4. *Optionally Credentialed* occupations are unregulated. However, closely related post-secondary programs exist. For example, chefs may study in college culinary programs and human resource managers may have college diplomas in human resources or university degrees in labour

---

5. Physicians are not entirely similar to lawyers and veterinarians in this regard. Medical schools provide graduate degrees that can be undertaken only after graduating from a post-secondary program. Law schools grant second-entry undergraduate degrees. Students cannot be admitted without having completed some post-secondary education, but some schools do admit students who have completed two or three years of undergraduate study and allow them to either continue both degrees simultaneously or abandon their first undergraduate program and complete only the professional degree. Despite granting doctoral-level degrees, veterinary students may also enter their professional programs without completing an undergraduate degree

relations. Not everyone working in these occupations need have pursued related FOSs in their post-secondary studies, but many will have done so.

5. *Certification-based* occupations are regulated occupations requiring licensing. However, this licensing process is unrelated to post-secondary education programs. Licensing for these occupations is based on testing, with preparation for tests consisting of independent study or workplace-based courses or training.

6. *Open* occupations are not closely related to any field of study. We include an example of one occupation typically requiring post-secondary study and two for which post-secondary study is not typically required.

Examining the measures shows that in general the expected pattern of occupational rankings holds. Graduate professions, direct entry professions, and trades, the three categories of occupations with required FOS show the strongest occupation/education links, followed by optionally credentialed occupations, certification-based occupations, and open occupations, in that order. Looking at individual measures shows that degree centralization and measures indicating the proportion of workers in the top FOS and top 3 FOSs perform mostly as expected. They show consistently strong links in three types of occupations, with the strength of the link dropping off in each subsequent category. Measures of the proportion of workers in the top 5 and top 10 FOS do not show as steep a drop between the first three categories and subsequent categories. The index of diversity and index of qualitative variation do rank some occupations in the first three categories of occupations as more tightly linked to FOS than the subsequent categories. However, closer examination shows that the strength of occupation/education link within these categories vary more than they do for other measures. Degree centralization shows the expected pattern of rankings.

**Table 3: Bivariate Correlations Between Common Measures Calculated from at least 100 Workers and Measures Calculated from a Randomly Selected Sample of Workers in the Occupation**

	50 Workers	25 Workers	10 Workers	5 Workers	N
<b>PropTop1</b>	0.917	0.929	0.674	0.591	128
<b>PropTop3</b>	0.882	0.899	0.672	0.466	128
<b>PropTop5</b>	0.875	0.871	0.593	-	128
<b>ID</b>	0.852	0.871	0.685	0.642	128
<b>IQV</b>	0.833	0.851	0.435	0.397	128
<b>Degree Centralization</b>	0.909	0.921	0.506	0.374	128

Note: All correlations are statistically significant  $p < .01$

*Testing the Robustness of Measures*

Each measure is calculated using data from survey respondents who are incumbent in those occupations. While respondent survey weights are used in calculating these measures, the *actual* number of respondents incumbent in each occupation places mathematical limits on the range within which the education/occupation link can vary. At the most extreme, an occupation held by only one survey respondent would be maximally linked to a single FOS for every measure except degree centralization, which could not be calculated. While the exact limits will vary with the weights attached to the specific respondents in each occupation, occupations employing fewer respondents will be more limited than those with more.

Because of the limits created by having small numbers of respondents in an occupation — and to comply with Statistics Canada's data disclosure regulations — our measures were calculated only for occupations with at least five unique respondents. However, without empirical examination, it is unclear how small the respondent pool can be before measures cease to be reliable. To evaluate the robustness of our measures to small within-occupation samples we simulate small within-occupation samples for all occupations held by at least 100 unique respondents and recalculate our measures from these simulated data. There are 128 occupations held by at least 100 unique respondents. From each we select random samples of 50, 25, 10, and 5 workers. Samples are independent of one another across and within occupations. We calculate each measure from these samples and correlate measures from each random sample with measures calculated from respondents in the occupation.

Table 3 shows bivariate correlations between measures calculated for occupations with at least 100 unique workers and smaller random samples of respondents from within those occupations. Results are similar for all measures: Correlations are quite high for samples of 50 and 25, but drop below .8 for samples of 10 and 5. The measures calculating degree centralization and the proportion of workers in the most common occupation are especially robust to modest-sized samples, though correlations for degree centralization drop more precipitously for very small samples than correlations for any other measure. As a whole these findings suggest that measures are robust for occupations with at least 25 workers, but measures based on smaller samples should be used with caution.

**Table 4: Regression Coefficients for Models Predicting Characteristics of Jobs Held in 1997**  
**Using Common**

	Program Intended to Prepare (Standard Error)	Employer Specified Field of Study (Standard Error)	Work is Related to Ed. Program (Standard Error)
PropPop1	2.647388*** (0.1258199)	1.40827*** (0.1533975)	0.9833061*** (0.0437742)
PropPop3	3.550377*** (0.1502951)	2.169713*** (0.2023642)	1.285247*** (0.0542223)
PropPop5	4.146928*** (0.1769658)	2.632769*** (0.2473116)	1.511971*** (0.0652637)
PropPop10	6.045201*** (0.2647858)	3.954042*** (0.3843808)	2.213432*** (0.0991367)
ID	2.765409*** (0.1411991)	1.539025*** (0.1610518)	1.022373*** (0.0452444)
IQV	2.704319*** (0.1416463)	1.466594*** (0.1605154)	1.002356*** (0.0448845)
Degree	2.594769*** (0.1232447)	1.378276*** (0.1499582)	0.9620749*** (0.0429517)
Centralization			
N	18964	12574	15800
Regression Type	Logit	Logit	OLSI

\* p<.1 \*\* p<.05 \*\*\* p<.01

Note: All regressions control for gender, respondents' highest level of education, and the number of respondents in the occupation.

### *Testing the Validity of the Measures*

Next we test the validity of these measures by using them to predict respondents' reports of how related their work is to their educational program. We use three outcome variables. The first is a dichotomous variable coded 1 if respondents answered yes to the question "Was the educational program you completed in 1995 intended to prepare you for this job?" and coded 0 if they responded that it was not. The second outcome

variable is a dichotomous variable coded 1 if respondents answered yes to the question “Did the employer specify that [the required level of education] must be in a specific field of study?” and coded 0 otherwise. The third is a variable based on responses when asked how closely their jobs were related to the diploma or degree completed in 1995. Respondents could respond that their jobs were “not at all related,” “somewhat related” or “closely related” to the program completed in 1995.

We use logit models for the first two outcomes and OLS<sup>6</sup> regression for the third. Because occupations are gender-labeled and people with higher degrees are more likely to work in jobs related to their fields, and because measures for smaller occupations may be less reliable, we control for gender, the respondents highest of education, and an ordinal variable indicating the number of respondents in the occupation (Boudarbat and Chernoff 2009, Reskin and Padovic 2002)<sup>7</sup>.

Results of each regression are shown in Table 4. Each cell in the table presents the regression coefficient of the measure, and the standard error of the co-efficient for a single regression model. Because all measures range from 0 to 1, co-efficients for models predicting the same outcome variable are interpreted similarly. Across all models and measures every co-efficient is statistically significant. Looking at the relative size of co-efficients shows that across all three outcome variables, the measures which were earlier shown to be less exclusively based on the popularity of the most common FOS are the better predictors of these outcome variables. The proportion of workers in the three and five most common FOS, consistently show the strongest associations. Of the measures influenced primarily by the proportion of workers in the most common FOS, degree centralization and the proportion of workers in the most common FOS are stronger predictors than ID or IQV of a respondents reports that they enrolled in a program to prepare them for their current job and that their employer required a specific field of study, though ID is slightly stronger in predicting whether respondents jobs are related to their 1995 degrees.

---

6. Ordinal regression showed substantively similar results. We present the OLS parameters because they are more intuitive to interpret.

7. We omit three occupations from these analyses: University teachers and professors, College and university teaching assistants and college and university research assistants. Because occupational codes do not distinguish between teachers and researchers in different subjects and because every field of study requires teachers and those teachers will likely have credentials in the subjects they teach, these three occupations were consistently ranked the three most weakly-linked occupations by all measures. We treat this as the result of coding not appropriate to our theoretical purposes and omit these occupations from the models.



## DISCUSSION

The measures of occupation/education link created here are based on the characteristics of incumbents in the occupation. This approach to measuring characteristics of occupational positions is common (e.g. Boyd 1986, Nam and Boyd 2004). In the case of occupations' link to FOSs, this approach allows us to determine the *de facto* requirements for entry and permeability of occupational boundaries rather than examining the skills or knowledge objectively required to perform a job as these can differ considerably and in unexpected ways. (Cain and Treiman 1981, Brown 1995).

Four measures vary primarily in the popularity of the most common FOS in the occupation. These are the proportion in the most common field of study, the ID, the IQV, and degree centralization. The remaining two measures are based on the proportion of workers in the three and five most common FOSs. Measures within these two classes consistently behave similarly to one another and differently from measures in the other class across the range of analyses we run. Therefore, we analyze the uses of our measures here separately for each class of measure and then discuss of which class of measure to use.

Among those measures that vary primarily with the single most common FOS, degree centralization and the proportion of workers in the most common FOS perform consistently better than ID and IQV. These measures are better able to rank sample occupations in theoretically expected ways, they are more robust to calculation using small samples of workers, and they are better predictors of outcomes that strong occupation/education link should predict. Though both measures perform similarly, we recommend the proportion of workers in the most common FOS from this group for two reasons. First, the measure is more intuitive to calculate, explain, and interpret than degree centralization. Second, though both measures perform reliably where there are at least 25 people in an occupation and less so for smaller occupations, the reliability of degree centralization drops considerably more steeply. Researchers wishing to use measures calculated from these smaller occupations should prefer the more robust measure.

The three measures based on the proportion of workers in the three, five, and ten most common FOS within an occupation allow researchers to consider occupations as tightly linked to FOSs when they have not only a single tightly linked FOS, but also if they have a small number of FOSs closely tied to the occupation. We discard the proportion of workers in the 10 most common FOSs because this measure does not differentiate well between occupations with strong occupation/education

links. As is apparent in the correlations shown in Table 2, this measure hits the ceiling value of 1 relatively early, with more than 25% of occupations receiving this same score. From the remaining two measures, the proportion of workers in the three most common FOSs performs best: In ranking the sample occupations, it distinguishes better between the very strongly linked to FOS categories of professions and trades and the less strongly linked optionally credentialed occupations; it is more robust to small samples, and by a very tiny margin, but a margin nonetheless, it shows stronger predictive validity. That said, in all cases the differences are small and thus the proportion of workers in the five most common FOSs remains a reasonable choice of measure, particularly for applications for which researchers may for theoretical reasons wish to consider the importance of a larger number of fields.

The choice of approaches — one FOS vs. multiple FOSs — is a theoretical one. Strictly speaking, an occupation that can be entered via only a single FOS is more tightly linked to educational field than an occupation with multiple paths to entry. However, this distinction is primarily important at the very strongly linked end of the occupational spectrum. At the centre of the continuum, single-FOS-based measures cannot distinguish between an occupation in which the most common FOS is held by 40% of workers and the remaining workers are spread evenly across 15 other FOSs, and an occupation in which the most common FOS is held by 40% of workers and the second most common is held by 35% of workers. Such a situation could result from an occupation in transition, competing claims over legitimacy of workers, historical contingencies creating two paths to entry, or methodological choices in the coding of closely-related FOSs. While most researchers would consider the latter more tightly linked, measures based on the single most common FOS will rank them similarly. Researchers dealing with data or research questions where such a situation is likely or where they are theoretically relevant, may prefer a multiple-FOS-based measure.

## CONCLUSION

This paper is a contribution to the literature and future research both in the sociology of occupations and on job-worker match and mismatch. To the sociology of occupations we offer a set of measures that can be combined with data sets using Statistics Canada occupational codes. These measures can be calculated for other occupational coding systems including those of other countries, where data on educational backgrounds is available; most frequently this will be from surveys of graduates (e.g.

NSF 2012) or longitudinal surveys following youth from school through labour market entry (e.g. Bureau of Labor Market Statistics 2005, Harris et. al. 2009). For researchers not requiring measures of occupations/education link across entire labour markets, these measures could, in principle, be calculated for individual occupations, where data are available. However, extreme caution should be exercised in calculating these measures from data that are not broadly representative of the focal occupations. For example, data from a single organization would not yield measures that could be meaningfully compared to those calculated here.

To the literature on job match we make the theoretical suggestion that understanding the consequences and causes of match and mismatch requires not only that we expand our view of educational match beyond education level to look at skills or areas of education, but that we consider the possibility that some occupations facilitate match and mismatch while for other occupations neither a strong match nor a strong mismatch is easily achieved. Measuring this possibility of match as a continuum allows researchers to distinguish between workers who report that their education is not strongly related to their work because they lack the appropriate and common background and those whose education is not strongly related to their work because no strongly related educational program exists.

## REFERENCES

- Abbot, Andrew. 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.
- Adams, Tracy L. 2000. *A Dentist and a Gentleman: Gender and the Rise of Dentistry in Ontario*. University of Toronto Press.
- Allen, Jim and Rolf van der Velden. 2001. Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers* 53: 434-452.
- Bills, David B. 2003. Credentials, signals, and screens: Explaining the relationship between schooling and job assignment. *Review of Educational Research* 73: 441-449.
- Borgatti, Stephen P. and Martin G. Everett. 1997. Network analysis of 2-mode data. *Social Networks* 19: 243-269.
- Bourdarbat, Brahim and Victor Chernoff. 2009. The determinants of education-job match among canadian university graduates. *IZA Discussion Paper* 4513: 30.

- Bourgeault, Lynn. 2000. Delivering the 'new' canadian midwifery: the impact on midwifery of integration into the ontario health care system. *Sociology of Health & Illness* 22: 172-196.
- Boyd, Monica. 1986. Socioeconomic indices and sexual inequality: A tale of scales. *Canadian Review of Sociology/Revue canadienne de sociologie* 23: 457-480.
- Brown, David K. 1995. *Degrees of Control: A Sociology of Educational Expansion and Occupational Credentialism*. New York: Teachers College Press
- Bureau of Labor Statistics, U.S. Department of Labor. *National Longitudinal Survey of Youth 1997 cohort, 1997-2003 (rounds 1-7)* [computer file]. Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2005.
- Cain, Pamela S., and Donald J. Treiman. 1981. The dictionary of occupational titles as a source of occupational data. *American Sociological Review*. 46: 3: 253-278.
- Carley, Kathleen and Matt DeReno. 2006. *ORA 2006: User's Guide*. Pittsburgh: Institute of Software Research, Carnegie Mellon University.
- Chillas, Shiona. 2010. Degrees of fit? Matching in the graduate labour market. *Employee Relations* 32: 156-170.
- Collins, R. 1979. *The Credential Society: An Historical Sociology of Education and Stratification*. New York: Academic Press.
- Everett, Martin and Stephen P. Borgatti. 2005. Extending centrality. in *Models and Methods in Social Network Analysis*, edited by P. Carrington, J. Scott, and S. Wasserman. Cambridge: Cambridge University Press.
- Freeman, Linton. 1979. Centrality in social networks: conceptual clarification. *Social Networks* 1: 215-39.
- Handel, Michael J. 2003. Skills mismatch in the labor market. *Annual Review of Sociology* 29: 135-165.
- Harris, K.M., C.T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J.R. Udry. 2009. *The National Longitudinal Study of Adolescent Health: Research Design*. <http://www.cpc.unc.edu/projects/addhealth/design>.
- Heijke, Hans, and Christoph Meng, and Catherine Ris. 2003. Fitting to the job: the role of generic and vocational competencies in adjustment and performance. *Labour Economics*. 10: 2: 215-229.
- Hersch, Joni. 1991. Education Match and job match. *The Review of Economics and Statistics*. 73: 1: 140-144.
- Kalleberg, Arne L. 2007. *The Mismatched Worker*. New York: W. W. Norton & Company.

- Layne, China. 2010. The demographics of over-education in the United States, 1971-2006. pp. 1-35 in *International Labour Process Conference*. New Brunswick, NJ.
- Liebertson, Stanley. 1969. Measuring population diversity. *American Sociological Review*. 34: 6: 850-862.
- Murphy, Raymond. 1984. The structure of closure: a critique and development of the theories of Weber, Collins, and Parkin. *The British Journal of Sociology* 35: 547-567.
- Nam, Charles B., and Monica Boyd. 2004. Occupational status in 2000: Over a century of census-based measurement. *Population Research and Policy Review*. 23: 327-358.
- National Science Foundation, Science Resource Studies Division. 2012. *National Survey of College Graduates (NSCG)*. Washington, DC: National Science Foundation
- Ortiz, Luis and Aleksander Kucel. 2012. Do Fields of study matter for over-education? The cases of Spain and Germany. *International Journal of Comparative Sociology*. 53: 305-327
- Reskin, Barbara and Irene Padavic. 2002. *Women and Men at Work*. Thousand Oaks: Sage.
- Robst, John. 2007. Education and job match: The relatedness of college major and work. *Economics of Education Review*. 26: 4: 397-407
- Robst, John. 2007. Education, college major, and job match: gender differences in reasons for mismatch. *Education Economics*. 15: 2: 159-175
- Robst, John. 2008. Overeducation and college major: expanding the definition of mismatch. *The Manchester School* 76: 349-368.
- Sloane, Peter J. 2002. Much Ado about nothing? What does the over-education literature really tell us? *Keynote Address International Conference on Over-education in Europe: What Do We Know?* Berlin: November
- Sørensen, Aage. 1979. A model and a metric for the analysis of the intragenerational status attainment process. *American Journal of Sociology*. 85: 361-384
- Statistics Canada. 1998. *User Guide - National Graduates Survey - Follow-up of 1995 Graduates*. Ottawa: Statistics Canada
- Weber, M, G Roth, and C Wittich. 1922 [1978]. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.
- Weeden, Kim A. 2002. Why Do some occupations pay more than others? Social closure and earnings inequality in the United States. *American Journal of Sociology* 108: 55-101.
- Van de Werfhorst, Herman G. 2002. Fields of study, acquired skills and the wage benefit from a matching job. *Acta Sociologica*. 45: 4: 287-303

- Van de Werfhorst, Herman G., and Robert Anderson. 2005. Social background, credential inflation and educational strategies. *Acta Sociologica*. 48: 4: 321-340
- Weiss, Richard M and Lynn E. Miller. 2010. The social transformation of American medical education: class, status, and party influences on occupational closure, 1902–1919. *Sociological Quarterly* 51: 550-575.
- Welsh, Sandy, Merrijoy Kelner, Beverly Wellman, and Heather Boon. 2004. Moving forward? Complementary and alternative practitioners seeking self-regulation. *Sociology of Health & Illness* 26: 216-241.

**Alexandra Marin** is an Assistant Professor of Sociology at the University of Toronto. Her research focuses on social networks, the flow of job information through social networks, including occupational variation, social network change, and methods for collecting social network data.

**E-Mail:** [alexandra.marin@utoronto.ca](mailto:alexandra.marin@utoronto.ca)

**Sean Hayes** is a PhD candidate in the Department of Sociology at University of Toronto. His research examines ethnicity and labour markets in early 20th century Canada.

**E-Mail:** [s.hayes@mail.utoronto.ca](mailto:s.hayes@mail.utoronto.ca)