

2020

36th IARIW General Conference

Paper Prepared for the 36th IARIW General Conference, Oslo, Norway, August 24-28, 2020

What's in a job? Measuring Skills from Job Vacancies

Gueorguie Vassilev

Khloe Evans

Vahe Nafilyan

Cognitive versus non-cognitive skills academic debates have inconclusively argued for which is more valuable for people's outcomes. For example, Heckman, Stixrud and Urzua (2006) infers equal importance from cognitive and non-cognitive skills, and with a low-level grouping of skills, is able to explain a large proportion of subsequent educational and labour market outcomes.

Generally, social survey sources suffer from measurement and definition errors, as well as biases in the self-reported measures of non-cognitive skills. Alternatively, they require extensive testing, such as OECD's PISA and PIAAC surveys, and are more robust in cognitive measures. This paper takes an alternative measurement approach, looking at which skills are asked for in online job vacancies.

Building on work from NESTA and Burning Glass, as well as more static sources of information such as O-Net and ESCO, we propose linking skills and other job requirements to jobs, as well as the advertised salaries. This provides a platform to infer the relative demand for different skills, whether cognitive or non-cognitive, and what are the interactions between them. Taking this further, it enables the quantification of such demand, allowing for an inferred 'demand price' of skills.

The method proposed will look to apply natural language processing techniques to extract relevant requirements per jobs. Subsequently, hedonic regression-type methods, such as applying machine learning techniques, should enable to model non-linear relationships between the requested skills and salary. Interesting linked skill groups will be discussed, as well as skills which have non-linear impacts onto others. Particular focus will be paid to other non-skill information relevant for determining proposed salaries, whether that is occupation, region, industry or other. Additionally, the appropriate level of categorisation will be discussed, such as the level of skill detail and if particularly non-cognitive skills can be grouped or whether they have distinct influences on salaries, as well as a consideration on missing or implied skills.

The drawbacks in such a type of data source will also be discussed, and what they may mean for subsequent survey collection.