아태연구 제20권 제2호(2013)

Linguistic Diversity, Unemployment and Schooling:
: Empirical Evidence from the East-Asian Countries

Alexandre Repkin*
Seung-Hoon Song**

| 목 차 |

| 논문요약 |

노동시장에서 언어의 다양성이 실업에 미치는 직접 및 간접효과를 탐색하였다. 연구를 위하여 기술상보성(skill complementarity), 자기민족중심적 행동(ethnocentric behavior), 사회적 응집성(social cohesion) 및 정치적 열망(political aspirations) 등에 관한 가설을 설정하고, 이들의 검증을 위하여 동남아시아 17개국의 언어표본을 바탕으로 언어상호간의 이해성을 고려하여 언어의 다양화를 측정하였다. 사회적 응집성 (social cohesion) 및 정치적 열망(political aspirations)에 관한 가설은 실업에 대한 언어다양성의 간접적인 효과로서, 이는 한편으로는 소수언어 집단의 사회적 융합을 촉진시키는 교육제도를 통하여 나타나며 다른 한편으로는 민족언어학의 차이로 인한 갈등의 위험을 증가시키는 효과로 나타나기도 한다. 가설검증 결과 직장에서

* First Author: Assistant Professor, Economics Department, Konkuk University.
** Corresponding Author: Associate Professor, International Trade Department, Konkuk University.

기술상보성(skill complementarity)의 효과를 명백히 보이는 초등학교 졸업자의 경우를 예외로 하고, 자기민족중심적 행동(ethnocentric behavior)과 정치적 야망(political aspirations)에 관한 가설은 확실하게 입증되었다.

▪ 주제어: 실업, 언어의 다양화, 교육, 기술상보성, 민족갈등

I. Introduction

The focus of this study is on the channels through which linguistic diversity and schooling are affecting the rate of unemployment in the East Asian countries. While the level of educational attainment has been shown to produce a significant impact on the labor market outcomes, most of the attention appears to have been focused on the link between schooling and earnings, e.g. Griliches (1977) or Card (2001). A smaller number of studies focused on the link between education and unemployment. Thus, Riddell and Song (2011) study the relationship between educational attainment and re-employment rates in the U.S. labor market and find that the probability of being re-employed is positively related to the level of education. However, the authors also note the mixed nature of evidence on the relationship between formal schooling and the unemployment rate. Nickell (1979) finds that more education increases the probability of finding a job, and Mincer (1991) also reports evidence of a lower risk of unemployment at higher levels of education. The general idea behind these studies is that the job search costs are lower, and the probability of meeting the employers' requirements are higher for job applicants commanding better skills, which are an increasing function of education.

The aforementioned studies, however, are overlooking an important aspect of the social environment in which the job-searching process is taking place, namely, the ethno-linguistic background of the job candidates. The crude idea is, employers will have both incentives and disincentives alike to hire a person whose mother tongue is different from their own. On the one hand, as is amply demonstrated by the role biodiversity plays in the survival of the ecosystems, a firm combining idiosyncratic skills pertaining to various ethno-linguistic backgrounds is likely to be producing more efficiently. On the other hand, politically motivated ethno-linguistic preferences may be preventing employers from making employment decisions on the basis of productivity considerations alone. One may choose, for instance, to hire workers whose mother tongue is the same with the employer's.

There have recently appeared studies that attempted to integrate the concept of ethnic and linguistic diversity into the economic analysis. Alesina and La Ferrara (2005) provide a survey of this literature that generally emphasizes a negative role to the ethnolinguistic diversity due to its inherent ability to create conflict potential. The contribution of this paper is twofold. First, we employ a refined version of the linguistic diversity variable that takes account of the extent to which languages are mutually intelligible (see e.g. Repkine 2013). Second, we relate unemployment rates both to linguistic diversity, and to the level of educational attainment. In this regard, we recognize the indirect effects through which linguistic diversity may affect unemployment. Thus, while more schooling may act as a social cohesion instrument by integrating various language groups into the mainstream society, it may also raise the sense of political awareness by the language and ethnic minorities, potentially resulting in an ethnicity-related conflict. We choose the East Asian region for our analysis since it is one of

the most linguistically diverse regions in the world, spanning countries from the very homogeneous Korea to countries as diverse as India and China.

This paper is organized as follows. In the next section we formulate the hypothesis on the links between linguistic diversity, educational attainment, and unemployment. Section III describes how linguistic diversity is measured, focusing on the possibility of languages' mutual intelligibility. Section IV describes the data, Section V discusses empirical results, and Section VI concludes.

II. Why Would Linguistic Diversity and Schooling Affect Unemployment?

The answer to this question lies in the fact that decisions related to employment are always made in the context of interaction between the employer and the prospective employee. In a multilingual, multicultural environment such interactions often involve the two parties negotiating an employment deal coming from different linguistic or cultural backgrounds. We believe there are at least two effects at work here. First, the employer is obviously looking for someone whose marginal product of labor is high enough to justify the wage, which can be the case if the skills of a job applicant are complementary to the skills of incumbent employees. In other words, linguistic diversity is associated with an increased scope for such skill complementarity that is conducive to productivity gains due to e.g. more specialization. This is one of the assumption in the theoretical model by Alesina and La Ferrara (2005), namely, that ethnically (and

hence, linguistically) diverse societies are more productive compared to their more homogeneous counterparts. A more diverse linguistic environment results in a broader pool of skills for the employer to choose from, resulting in a higher probability of finding a job, and lower unemployment rates as a result. We call this the skill complementarity effect of linguistic diversity.

Second, linguistic diversity undoubtedly entails the need to overcome some sort of communication barriers. We are not talking solely about the inability to communicate in a common language. It may be that both negotiators are equally fluent in a local lingua franca, while possessing conflicting mentalities due to their ethnic or linguistic identity. In this context, it does not really matter whether a native speaker of language A is fluent or not in language B, the language of his negotiating party. It is the fact that language A speaker has a different mentality form the language B speaker that is of importance to the success of the employment negotiation. In this way linguistic differences are representing barriers to negotiations that sometimes can be high enough to result in the ethnocentric behavior. For instance, employers might choose to hire only the candidates speaking a particular language (e.g. their own) irrespectively of their objective productivity. We call this the ethnocentric behavior effect, which is a well-known phenomenon in the literature. Thus, Easterly (2001) is modeling conflict incidence as an increasing function of ethnic diversity. Mauro (1995) finds that ethnicity and language-related rent-seeking behavior results in corruption and general inefficiency in economic decision-making.

In this study we are testing whether it is the skill complementarity, or the ethnocentric behavior effect that is taking over when it comes to employment decisions. However, we also recognize the fact that ethnic diversity can affect unemployment in an indirect fashion as

well, in particular through the educational system. Since primary and secondary schools in Asia are not organized along a linguistic principle, entering a school often means a chance to integrate more with the rest of the society, underscoring the role of schools as a social cohesion instrument. Higher school enrollment rates in this case will mean a wider scope for exploiting the already mentioned skills complementarities in the future. In case schooling is not widely available, the complementarity of skills due to diverse linguistic backgrounds may fail getting translated into improved labor market conditions for a variety of reasons ranging from the inability of the members of various linguistic groups to communicate with each other, and to the lack of experience communicating in a multicultural environment. On the contrary, wide availability of schooling accomplishes the important task of social integration of pupils with different linguistic and cultural backgrounds into the mainstream society by encouraging communication in a local lingua franca, and by nurturing useful cross-cultural communication skills.

More schooling, however, may also raise political awareness of the pupils with respect to either current or historical political issues that might prove an obstacle on the way of their integration with the mainstream society. This is especially relevant for the case of ethnic and linguistic minorities that might grow increasingly aware of the discrimination issues both as a result of more learning about the historical past at school, or directly as a consequence of possible mistreatment by another (very possibly dominating) ethno-linguistic majority in class. In this case, a lot depends on the quality of education. For these reasons we are considering both the quantity and the quality of education in our study to account for the indirect effects of linguistic diversity on unemployment.

To summarize, in this study we are aiming to test the following

four hypotheses pertaining to the effects of linguistic diversity on the labor market conditions, in particular on the levels of unemployment:

1. *Skill complementarity*. A linguistically diverse environment provides wider scope for the exploitation of complementarities of skill pertaining to workers with different ethno-linguistic backgrounds, which increases their productivity and in this way makes them more attractive for hire by the prospective employers.
2. *Ethnocentric behavior*. Linguistic diversity might become source of social cleavages resulting in some employers being reluctant to hire workers with an ethnolinguistic background other than their own, producing a negative influence on the labor market in general, and unemployment rates in particular.
3. *Social cohesion*. A greater extent of educational availability encourages integration by linguistic minority groups into the mainstream society by stimulating the use of a common language and by nurturing skills of communicating in a multicultural environment. A greater extent of social integration in turn results in more opportunities to find employment so that the effects of ethnic diversity on the unemployment rates will be different depending on the quantity of education.
4. *Political awareness*. More schooling may equally result in more awareness of the ethnocentric behavior by the ruling ethnolinguistic majority either as a result of the knowledge (e.g. about historical past) acquired in the process of education, or as a consequence of direct interacting with the dominant ethnolinguistic group at school. As a result, employment decisions in the future may be tainted by linguistic preferences, representing a negative effect linguistic diversity might produce on the unemployment rates through more schooling. Political awareness resulting from more education in this way is reinforcing the ethnocentric behavior effect.

III. Measuring Linguistic Diversity

The problem of measuring the extent of linguistic diversity was apparently first tackled by Greenberg (1956) who designed a set of related measures, the simplest one being the monolingual nonweighted index computed as follows:

$$EDI = 1 - \sum_{i=1}^{n} s_i^2 \qquad (1)$$

where s_i are the shares of a country's population speaking language $i = 1..n$, and n is the total number of languages. EDI can be interpreted as the probability of any two randomly chosen persons of not being able to communicate. As the extent of linguistic fragmentation grows, EDI approaches unity. One implicit assumption behind the index in (1) is that every person is monolingual, i.e. is able to communicate in only one language. Another assumption behind (1) is that any two languages in a country are mutually unintelligible, which is clearly counterfactual. To allow for the possibility of mutual intelligibility between languages, Greenberg (1956) suggested a weighted version of the language diversity index, EDIW, computed as

$$EDIW = 1 - \sum_{i=1}^{n} \sum_{j=1}^{n} s_i s_j r_{ij} \qquad (2)$$

where r_{ij} is the extent to which the two languages i and j are mutually intelligible.

Dryer and Haspelmath (2011) compiled the WALS[1] database that for each language lists an array of more than one hundred

1) WALS stands for World Atlas of Language Structures.

characteristics related to syntax, grammar, phonetics and other linguistic features. We employ the approach by Lohmann (2011) to compute the mutual intelligibility index $r_{ij} \in [0,1]$ as a share of linguistic features with the identical score for languages i and j. The larger this share, the greater the extent to which the two languages are mutually intelligible. Clearly, any language is perfectly intelligible with itself, i.e. $r_{ii} = 1$. In case $r_{ij} = 0$ in case $i \neq j$, the two languages i and j are mutually unintelligible, and EDIW=EDI. If $r_{ij} \neq 0$ reduces, $EDIW \leq EDI$ which is another way to say that in any country the probability of two randomly chosen people not to be able to communicate with each other is always lower than or equal to the same probability computed for the same country where all languages are completely different. We can say then that the weighted measure of linguistic diversity EDIW is a more conservative version of the unweighted EDI.

The Ethnologue (Lewis, 2009) database provides numbers of speakers by language in each country. Zero mutual intelligibility is assumed (i.e. $r_{ij} = 0$) in case the WALS database did not contain linguistic characteristics of the languages reported by Ethnologue. Table1 presents summary statistics for the two indices of linguistic diversity in the sample countries.

Table 1. Indices of Linguistic Diversity

Mean	EDIW, %	EDI, %
Bangladesh	0.66	1.64
Bhutan	9.77	43.75
Brunei	36.05	82.55
Cambodia	1.65	10.53
China	10.67	36.14
India	39.33	91.70
Indonesia	49.80	75.61

Japan	0.31	1.12
Korea	0.00	0.00
Malaysia	63.64	80.57
Mongolia	6.05	28.55
Myanmar	20.26	40.72
Pakistan	45.97	73.96
Philippines	34.87	80.26
Singapore	37.88	71.51
Thailand	6.39	17.44
Vietnam	8.29	18.12

As demonstrated by Table 1, allowing for the languages' mutual intelligibility sometimes greatly reduces the value of the linguistic diversity index. Thus, in case of India the value of the index decreases by more than one-half, while in China the decrease is almost four-fold, and it is a six-fold decrease in the case of Cambodia. Korea is conspicuous in that both weighted and unweighted values of its linguistic diversity index are equal to zero, which is not surprising given the fact that Korea is one of the world's most homogeneous countries in both ethnic and linguistic sense[2]. Japan is similarly estimated to be linguistically very homogeneous. Except for these cases, allowing for the mutual intelligibility between languages appears to be greatly reducing the measured extent of linguistic diversity in the region's countries.

While we are assuming that the extent of linguistic diversity is a time-invariant characteristic, some researchers such as Fedderke, Luiz and De Kadt (2008) argue that the extent of linguistic diversity may be changing with time, like it has been in South Africa, reflecting

2) It is true that there are dialects in Korea, but except for the dialect of the Jeju island located to the South of the Korean peninsula, all of these dialects are mutually intelligible to a large extent. While the Jeju dialect is not easily understood by most Koreans, most linguists agree it is a dialect of Korean rather than an independent language.

changes in the institutional and economic environment. However, we believe the extent of linguistic diversity in the East Asian region is not likely to have changed significantly during the ten-year period covered by our sample.

IV. Data Sources and Economic Variables

In our empirical work we use observations on seventeen Eastern, South-Eastern, and Southern Asian economies, dataset of which includes cross-section data of 17 countries and 10 year time-series data for the period of 2002-2011. We employ the World Bank's (2013) World Development Indicators (WDI) database for all variables employed in this study except for the linguistic diversity indices whose construction was described in Section III. In this study we use three measures of unemployment: the conventional unemployment rate, and the two shares of the latter that correspond to the contributions by the job seekers with only primary, or secondary school educational attainment. These are coded in the WDI as SL.UEM.TOTL.ZS, SL.UEM.PRIM.ZS and SL.UEM.SECO.ZS, respectively.

We measure the quantity of primary education PRIMQN and that of secondary education SECQN as the gross enrollment rate measured in percentage terms. We opted for the gross, rather than the net, enrollment rate since the former accounts for grade repetition, which is important in the context of this study. The WDI codes for these two variables are SE.PRM.ENRR and SE.SEC.ENRR, respectively. The choice of variables to measure the quality of primary and secondary education was hindered by both the data availability problem, and the

fact that for some measures of educational quality we were unable to find a suitable set of instruments that would ensure uniform comparison of the results across empirical specifications. As a result, we are using two different measures of educational quality for the primary and secondary education cases. The quality of primary education, PRIMQL, is measured as the share of trained teachers, who obtained official teacher's certificate from the Ministry of Education in their own country, in the pool of all teachers, and is coded by the WDI as SE.PRM.TCAQ.ZS. The quality of secondary education, SECQL, is coded as SE.SEC.REPT.ZS, and is measured as the share of grade repeaters in total enrollment.

We are also employing a number of instruments for our estimations. These are variables representing the labor market characteristics such as e.g. employment to population ratio, public spending on education, the extent of development of the telecommunications sector, the state of health care in the country, the general economic indicators, extent of economic inequality, and the structure of population. While the original list of candidate instruments consisted of fifty-five variables, some of them were never chosen such as e.g. the Gini index or in fact (rather surprisingly!) any variable in the group of economic inequality indicators. As a result the total number of instruments employed in the final set of our regressions amounts to eleven. These were chosen by an algorithm following the selection procedure described in Section V and encoded in STATA. The instruments' names are detailed in Tables 3 and 4 in Section V below, and a detailed summary is available from the authors upon first request. Table 2 below provides summary statistics for the core variables employed in our study.

Table 2. Summary Statistics

	Mean	SD	Min	Max
Unemployment, %	4.42	2.56	0.10	11.90
Unemployment Primary, %	32.68	18.87	7.80	84.70
Unemployment Secondary, %	38.77	17.80	3.00	71.90
Weighted Linguistic Diversity, %	21.86	20.04	0	63.64
GDP per capita, constant 2000 USD	7296	11554	328	40837
PRIMQN, %	106.45	11.33	72.10	130.20
PRIMQL, %	89.53	12.86	51.43	100
SECQN, %	69.35	22.14	22.66	111.84
SECQL, %	2.93	2.86	0	11.61

Note: unemployment secondary is measured as the share of total unemployment accounted for by the primary school graduates, unemployment secondary is measured analogously.

The total unemployment rate in the sample region is a decent 4%, with contributions from the primary and secondary school graduates accounting for 68% of the total unemployment pool. It is interesting that on average it is the secondary school graduates that are contributing six percentage points more to the total unemployment compared to their primary school counterparts. This is reflecting what we believe an important finding in this study: more education may reduce one's chances of landing a job since along with fostering skills that allow one to compete in the labor market, this same education is increasing the extent of the competition itself. As a result, it may be more difficult for a better educated person to find a job compared to someone with only basic schooling.

Enrollment rates in the primary and secondary schools sometimes exceed 100% in our sample due to the fact that, as we mentioned already, we are using the gross enrollment rates that account for grade repetition. Expectedly, the secondary school enrollment rates are lower than those for the primary school. One should keep in mind

that the variable measuring the quality of secondary school education, SECQL, is actually increasing with the deterioration of quality since it is equal to the share of grade repeaters in the total secondary school enrollment.

V. Empirical results

In our empirical work we model the rate of unemployment to be a function of linguistic diversity, and variables characterizing the educational environment that may be linked to the extent of linguistic diversity as well. As indicated by Fedderke and Luiz (2008), it is important to make a distinction between the educational quantity and quality since the two are not equivalent to each other. Also the level of schooling is important since arguably, higher levels of educational attainment are associated with higher chances of landing a job after graduation. Our benchmark empirical model for the unemployment rate U can be written down as follows:

$$U = U\left(EDIW, \left[PRIMQN, PRIMQL, SECQN, SECQL\right]\right) \qquad (3)$$

where PRIMQN and SECQN stand for the quantity, and PRIMQL and SECQL stand for the quality of primary and secondary education, respectively. EDIW is the weighted index of linguistic diversity in (2).

It is unlikely that the determinants of unemployment in (uform) be completely exogenous. Thus, higher chances of finding a job (i.e. lower unemployment levels) in the country may serve as an impetus for the youngsters to strive at getting a better education. In addition, both educational quantity and quality may depend on the level of linguistic diversity since e.g. lower chances of getting a job due to linguistic preferences by the potential employers may translate into disincentives to enroll in school.

In order to resolve the potential endogeneity problem in (3), we employ a two-stage instrumental variables approach along the lines of Cinyabuguma and Putterman (2011):

$$\begin{cases} U_{it} = \alpha_0 + \alpha_1 \vec{Y}_{it} + \alpha_2 \vec{X}_{it} + \varepsilon_{it} \\ \vec{Y}_{it} = \beta_0 + \beta_1 \vec{Z}_{it} + \beta_2 \vec{X}_{it} + \eta_{it} \end{cases} \quad (4)$$

where U_{it} is the unemployment rate in country i and time t, $\vec{Y}_{it}$ is the vector of educational environment determinants in (uform), $\vec{X}_{it}$ is the vector of exogenous controls, and $\vec{Z}_{it}$ is the vector of instruments. The instrumental variables in $\vec{Z}_{it}$ are uncorrelated either with ε_{it} or with U_{it}, but they are correlated with the potentially endogenous educational determinants $\vec{Y}_{it}$: $E(Z_{it}\varepsilon_{it}) = 0, E(Z_{it}U_{it}) = 0$, and $E(Z_{it}U_{it}) \neq 0$ for each component of the instruments' vector $\vec{Z}_{it}$. The vector of exogenous controls $\vec{X}_{it}$ includes the log-GDP per capita and the log of linguistic diversity.

In choosing our instruments, we employed the Wu-Hausman exogeneity test, and Sargan's (1958) test of overidentifying restrictions. Cragg and Donald's (1993) was then performed in order to decide on whether the LIML (limited information maximum likelihood) or 2SLS (two-stage least squares) estimation method is more appropriate. The choice of our instruments was inspired by Lohmann (2011).Except for a single case, the Cragg and Donald's test strongly favored the LIML estimation. Table 3 reports the results for the primary school educational quantity and quality.

Those with only primary education appear to be contributing less to the total unemployment rate in case they come from linguistically

diverse societies, as evidenced by the estimates in columns (a) and (b). However, as suggested by the estimates in (d), the total unemployment rate is likely to be higher in the linguistically diverse economies. In terms of our hypotheses, this evidence is supporting the complementarity of skills hypothesis in case of the job seekers with primary education only: diverse linguistic environment creates complementarity of skills among job seekers which makes it easier for them to find a job. However, since we do not observe the same type of association between linguistic diversity and lower total unemployment, we conclude that the positive effects of such diversity are limited to the pool of job seekers with primary education only. In contrast, in the overall unemployment pool it is the ethnocentric behavior that appears to dominate the labor market landscape. This conjecture is further corroborated if we look at the estimation results for the secondary school educational variables in Table 4. Indeed, in contrast to the primary school case, we find more linguistic diversity to be consistently associated with the higher levels of both total unemployment, and unemployment contribution shares by the secondary school graduates.

Regarding the educational quality, our results differ whether we look at the effect on total unemployment rate, or that on the contribution to the latter by the jobseekers with a particular educational attainment. In interpreting the results, it is important to keep in mind that the educational quality variable in Table 4, lnSECQL, is actually decreasing with quality since it is measured as the percentage of grade repeaters in secondary school. Therefore, higher values of lnSECQL are associated with the lower quality of secondary education. Keeping that in mind, it is easy to see that the quality of both primary and secondary education is associated with lower rates of the total unemployment, at the same time being associated positively with

the contribution of jobseekers with a particular educational background to the total unemployment pool. The former finding is rather intuitive since one would expect better educational quality to reduce the chances of not being able to land a job. The latter finding, however, cannot be readily explained either by conventional wisdom, or by the mainstream economics literature. Our explanation is that a better quality primary education, for instance, while increasing the qualifications of potential job seekers, is also increasing the strength of competition between the primary school graduates who enter the labor market after finishing their primary school studies. That competition has a potential to reduce the chances of getting employment within the pool of primary school graduates, while these graduates are not competing with those who have secondary school education since the primary school graduates do not qualify for the jobs available to the high school graduates. Similarly, better quality secondary schools are increasing the skills of their graduates, while also spurring more competition among those graduates, which in turns increases the size of their contribution to the overall unemployment pool. At the same time, the competition with primary school graduates is precluded for at least three reasons. First, as mentioned already, the primary school graduates do not possess skills necessary to get most, if not all, of the jobs offered to the secondary school graduates. Second, the secondary school graduates will likely self-select out of the jobs offered to the primary school graduates since these jobs typically offer lower pay and worse employment conditions compared to jobs that require a secondary education degree. Third, in many cases the menial jobs requiring only basic education are not accessible to the better educated job seekers on the grounds of the latter being "overqualified". This lack of competition between secondary and primary school graduates in the labor market is explaining the fact

that, while the educational quality is increasing the chances to be unemployed within the group of job seekers with a similar educational level, in the overall labor market a better quality education is conducive to lower overall unemployment rates.

It is worthwhile looking from the linguistic diversity perspective at the effects of the quantity of primary and secondary education on unemployment both in terms of the overall rates, and in terms of share contributions by the particular educational group. As evidenced by Tables 3 and 4, the effects of educational quantity are uniformly positive except for regression (c) in Table 4 where the coefficient for quantity of secondary education comes out statistically insignificant. This finding is interesting since it suggests "too much education" is decreasing one's chances to find a job irrespectively of the attained educational level. Since this effect does not have the quality dimension incorporated in it, we interpret our findings to suggest that more education (e.g. in terms of hours at school) is providing students with basic socializing skills that are increasing their ability, but at the same time increasing the extent of their competition with each other in the job market irrespectively of the level of their educational attainment. Thus, while more schooling increases the extent of social inclusion, it also expands the pool of potential job seekers, which naturally results in the intensified competition among them for the jobs.

Overall, our empirical results suggest that, when attempting to explain unemployment rates, it is important to take account of the two important dimensions that have so far failed to receive much attention in the literature. First, the extent of linguistic diversity is likely to produce an effect on one's chances of finding a job for the two conflicting reasons. On productive efficiency grounds, a more diverse society is more likely to exploit skill complementarities pertaining to

workers with idiosyncratic linguistic backgrounds that get translated into idiosyncrasies in productivity, management practices and the like. However, there is a counteracting political dimension of ethnic diversity that has to do with the possibility of ethnocentric behavior on the part of the employers who e.g. might prefer hiring people of a specific ethnicity without giving much heed to those people's productivity. According to our estimates, the skills complementarity effect outweighs the political ethnocentric effect in case of the primary school graduates, while the ethnocentric behavior starts dominating in case of the secondary school graduates. Indeed, while the function of primary education is to provide students with a set of basic skills like reading and writing, there appears to be little scope for politically motivated cleavages along ethnic and linguistic lines. However, as the educational level increases, political awareness of the minority linguistic groups is likely to increase as well, offsetting the positive effects of skills complementarities and spurring the negative influence of the ethnocentric behavior instead, which is corroborated by our empirical findings.

Second, better quality schooling has a potential to decrease one's ability to find a job within a group of one's peers due to increased competition among higher quality candidates, while at the same time increasing one's chances to find a job in the overall sense. To draw a simple example, better-quality secondary education compared to the primary one in country A will mean that the secondary school graduates find it harder to find a job compared to the primary school graduates. At the same time, it might be easier to find a job for anyone in country A compared to country B since the quality of education (both primary and secondary) is higher in A versus B.

Table 3: Unemployment and Linguistic Diversity in the Primary School

	(a)		(b)		(c)		(d)	
Estimation Method	2SLS		LIML		LIML		LIML	
Dependent Variable	Primary Education Quantity, gross enrollment	Unemployment, % of total	Primary Education Quality, share of trained teachers	Unemployment, % of total	Primary Education Quantity, gross enrollment	Unemployment, total	Primary Education Quality, share of trained teachers	Unemployment, total
Constant	−21.864*** (0.000)	84.619 (0.080)	−57.483** (0.002)	37.836 (0.142)	−24.05** (0.001)	146.315** (0.035)	9.659 (0.736)	296.382*** (0.002)
ln *EDIW*	0.035*** (0.000)	−0.308*** (0.000)	0.148*** (0.000)	−0.254*** (0.003)	−0.005 (0.823)	0.433*** (0.000)	0.090** (0.012)	0.264*** (0.001)
GDP per capita, log	0.084*** (0.000)	−0.130** (0.050)	0.009 (0.719)	0.096*** (0.003)	0.037** (0.010)	0.437*** (0.000)	−0.042 (0.406)	0.072 (0.385)
Year index	0.014*** (0.000)	−0.049** (0.044)	0.033** (0.001)	−0.022* (0.100)	0.014*** (0.000)	−0.080** (0.024)	−0.002 (0.870)	−0.142** (0.003)
ln *PRIMQN*		3.853*** (0.000)				2.884*** (0.007)		
ln *PRIMQL*				1.819** (0.036)				−2.405*** (0.006)

Telephone lines	0.029** (0.033)		−0.035 (0.087)	0.013 (0.848)
Employment to population ratio	0.374*** (0.000)			
Mortality rate, under-5	0.046 (0.130)			
Improved water source (% access)	−0.786*** (0.000)	−1.162*** (0.000)	−0.287 (0.081)	
Public spending on education, % of GDP		0.213*** (0.003)		−0.124 (0.479)
Primary school starting age (years)		0.105 (0.649)	0.600*** (0.000)	
Secondary school starting age (years)		0.704** (0.041)		
Health expenditure, public (% of GDP)			−0.046 (0.135)	0.269** (0.012)
Population ages 0-14 (% of total)			−0.048 (0.650)	
Wu-Hausman	40.25*** (0.000)	17.17** (0.026)	11.15*** (0.002)	9.54** (0.021)

Sargan	5.01 (0.171)		4.71 (0.194)		6.88 (0.142)		1.78 (0.410)	
Adj. R^2	0.63	0.51	0.90	0.98	0.60	0.34	0.36	0.80
Partial R^2	0.66		0.93		0.64		0.50	
Min. eigenvalue statistic	27.72		3.93		19.95		2.57	
2SLS relative bias at 5%	16.85		10.27		18.37		13.91	
2SLS nominal 5% Wald test	24.58		13.96		26.87		22.30	
LIML nominal 5% Wald test	5.44		3.87		4.84		6.46	

Note: ***, ** and * correspond to 1%, 5% and 10% significance levels, respectively; p-values in parentheses. Adjusted R^2 refers to both first and second-stage estimations. LIML 2-stage estimation employed in case the 2SLS nominal 5% Wald test statistic exceeds the minimum eigenvalue statistic at the 5% significance level. P-values reported in parentheses for Sargan score. Number of observations: 89 on average.

Table 4: Unemployment and Linguistic Diversity in the Secondary School

	(a)		(b)		(c)		(d)	
Estimation Method	LIML		LIML		LIML		LIML	
Dependent Variable	Secondary Education Quantity, gross enrollment	Unemployment, % of total	Secondary Education Quality, grade repetition	Unemployment, % of total	Secondary Education Quantity, gross enrollment	Unemployment, total	Secondary Education Quality, grade repetition	Unemployment, total
Constant	−15.000 (0.448)	143.36*** (0.000)	6.260 (0.953)	39.506 (0.394)	−15.393 (0.287)	166.540** (0.017)	−47.539 (0.747)	182.486** (0.037)
ln *EDIW*	−0.008 (0.701)	0.199*** (0.000)	0.283* (0.098)	0.112** (0.038)	−0.044** (0.015)	0.195*** (0.001)	0.146 (0.454)	0.479*** (0.000)
GDP per capita, log	−0.032 (0.479)	0.173** (0.010)	1.072*** (0.000)	−0.033 (0.704)	−0.092*** (0.002)	0.260** (0.019)	0.617* (0.069)	−0.232* (0.065)
Year index	0.007 (0.503)	−0.073*** (0.000)	−0.012 (0.828)	−0.018 (0.446)	0.007 (0.299)	−0.082** (0.019)	0.010 (0.897)	−0.089** (0.043)
ln *SECQN*		1.181*** (0.000)				−0.760 (0.223)		
ln *SECQL*				−0.160*** (0.000)				0.241** (0.026)
Telephone lines	0.123 (0.167)		−1.076*** (0.002)		0.162*** (0.000)		−1.434*** (0.002)	
Employment to population			2.970** (0.041)					

Improved water source (% access)	0.347 (0.318)						6.276*** (0.000)	
Public spending on education, total (% of GDP)							−0.890 (0.270)	
Secondary school starting age (years)	1.785*** (0.000)				1.497*** (0.000)			
Health expenditure, public (% of GDP)			−0.804** (0.025)		0.036 (0.556)			
Population ages 0–14, %	−0.348* (0.071)							
Improved sanitation facilities, % access	0.303*** (0.006)				0.291*** (0.000)			
Wu-Hausman	4.50** (0.04)		5.06** (0.032)		6.00** (0.017)		4.59** (0.040)	
Sargan	0.51 (0.47)		4.79 (0.091)		7.52 (0.057)		5.85 (0.054)	
Adj. R^2	0.66	0.79	0.49	0.31	0.71	0.10	0.53	0.44

Partial R^2	0.70	0.52	0.73	0.59
Min. eigenvalue statistic	14.52	15.87	14.37	16.82
2SLS relative bias at 5%	18.37	13.91	16.85	13.91
2SLS nominal 5% Wald test	26.87	22.30	24.58	22.30
LIML nominal 5% Wald test	4.84	6.46	5.44	6.46

Note: ***, ** and * correspond to 0.5%, 1% and 5% significance levels, respectively; p-values in parentheses. Adjusted R^2 refers to both first and second-stage estimations. LIML 2-stage estimation employed in case the 2SLS nominal 5% Wald test statistic exceeds the minimum eigenvalue statistic at the 5% significance level. P-values reported in parentheses for Sargan score. Number of observations: 89 on average

VI. Conclusions

In this study we are attempting to draw attention to the important role the linguistic background of job seekers plays in their chances to find a job. We employed a sample of seventeen East and South Asian countries to test several hypotheses related to the direct and indirect effects of linguistic diversity on the rates of unemployment in the region. The direct effects of linguistic diversity on unemployment are represented by the skill complementarity and ethnocentric behavior effect. A linguistically diverse society is capable of combining the diverse skills possessed by job seekers with different linguistic backgrounds in a greater variety of productive ways compared to a linguistically homogeneous country, which results in lower unemployment rates, which we call the skill complementarity effect. However, the same linguistic diversity can be detrimental to the employment practices due to the ethnocentric trends in employment-making decisions whereby the employers have preferences for the employees whose linguistic and ethnic background is not very different from theirs.

The indirect effects of linguistic diversity on unemployment are captured by the social cohesion and political awareness effects that both work through the educational system. On the one hand, available schooling increases the extent of social and political integration of linguistic minorities into the mainstream society, encouraging them to learn and communicate in the country's lingua franca, and by training the future job seekers to communicate and cooperate in a multicultural environment. At the same time, the same educational system may provide pupils with knowledge of the ongoing or historical oppression of the linguistic minority they belong to, possibly also exposing them to an unfavorable treatment by the dominating language group in

class. The resulting political awareness of linguistic minority groups may result in reluctance to work for or to hire someone with a different ethnolinguistic background than their own, which would reinforce the effect of ethnocentric behavior and increase the rate of unemployment.

We use the instrumental variables estimation technique to account for the possible endogeneity of the quantity and quality of education with respect to the unemployment rates. We find that, except for the primary education, linguistically more diverse societies are also the ones with higher rates of unemployment, which is evidence corroborating the ethnocentric behavior and political awareness hypotheses. The social cohesion hypothesis finds support if our analysis is limited to the primary school education only. The prevalence of the ethnocentric behavior effect with the only exception of primary education is also lending support to the political awareness hypothesis since it is likely that both the amount of knowledge and the extent of possible linguistic origin-based mistreatment will be more pronounced at the more advanced educational stages.

In addition, we discover that, while expectedly the better educational quality is associated with lower total unemployment levels, they are also positively related to the higher shares of total unemployment pool accounted for by the primary or secondary school graduates. We infer that, while better quality schooling increases the future job seekers' competitiveness, it also increases the extent of competition between these job seekers, which reduces their chances to find a job within the group of one's peers, while increasing those chances when computed across all of the job seekers.

The analysis in this paper would benefit from several improvements. First, since the two of our hypotheses, namely, those of the ethnocentric behavior and of political awareness, are hinging on the

ability of linguistic diversity to create potential for social conflict, it would be useful to incorporate the data on the incidence of such conflicts in the region. Second, linguistic diversity is not necessarily the only defining factor behind social cleavages, the other such factor being e.g. religious affiliation. Finally, the implicit assumption behind the analysis in this study is that linguistic and ethnic affiliations are equivalent to each other, while this may not necessarily be so. For instance, a considerable part of Malaysian Chinese have adopted the Malay language and customs, yet they are habitually considered to be of 'Chinese' ethnicity in the country, see e.g. Hirschman (1987). We leave exploration of the abovementioned issues to our future work.

| 참고문헌 |

Alesina, A. and E. La Ferrara(2005). "Ethnic Diversity and Economic Performance." *Journal of Economic Literature*. Vol. 43, pp. 762-800.

Card, David(2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems." *Econometrica*. Vol. 69, pp. 1127-1160.

Cinyabuguma, M. M. and L. Putterman(2011). "Sub-Saharan Growth Surprises: Being Heterogeneous, Inland and Close to the Equator Does Not Slow Growth Within Africa." *Journal of African Economies*. Vol. 20. No. 2, pp. 217-262.

Cragg, J. G. and S. G. Donald(1993). "Testing Identifiability and Specification in Instrumental Variable Models." *Econometric Theory*. Vol. 9, pp. 222-240.

Dryer, M. S. and M. Haspelmath(ed.)(2011). The World Atlas of Language Structures Online. Munich: Max Planck Digital Library. Available online at http://wals.info. (Accessed on June 15, 2013)

Easterly, W.(2001). "Can Institutions Resolve Ethnic conflict?" *Economic Development and Cultural Change*. Vol. 49. No. 4, pp. 687-706.

Fedderke, Johannes W. and John M. Luiz(2008). "Does Human Capital Generate Social and Institutional Capital? Exploring Evidence From South African Time Series Data." *Oxford Economic Papers*. Vol. 60. No. 4, pp. 649-682.

Fedderke, Johannes W., John M. Luiz and Raphael De Kadt(2008). "Using Fractionalization Indexes: Deriving Methodological Principles for Growth Studies from Time Series Evidence." *Social Indicators Research*. Vol. 85. No. 2, pp. 257-278.

Greenberg, J. H.(1956). “The Measurement of Linguistic Diversity.” *Language*. Vol. 32 . No. 1, pp. 109-115.

Griliches, Zvi.(1977). “Estimating the Returns to Schooling: Some Econometric Problems.” *Econometrica*. Vol. 45, pp. 1-22

Hirschman, Ch.(1987). “The Meaning and Measurement of Ethnicity in Malaysia: an Analysis of Census Classifications.” *Journal of Asian Studies*. Vol. 46. No. 3, pp. 555-582.

Lewis, M. P.(ed.)(2009). Ethnologue: Languages of the World, Sixteenth edition. Dallas, Texas: USA, SIL International. Available online at http://www.ethnologue.com. (Accessed on June 23, 2013)

Lohmann, J.(2011). “Do Language Barriers Affect Trade?” *Economics Letters*. Vol. 110. No. 2, pp. 159-162.

Mauro, P.(1995). “Corruption and Growth.” *Quarterly Journal of Economics*. Vol. 110. No. 3, pp. 681-712.

Mincer, J.(1991). “Education and Unemployment.” NBER Working Paper. No. 3838.

Nickell, S.(1979). “Education and the Lifetime Pattern of Unemployment.” *The Journal of Political Economy*. Vol. 87. No. 5, pp. 117-131.

Repkine, A.(2013). “Ethnic Diversity, Political Stability and Productive Efficiency: Empirical Evidence from the African Countries.” forthcoming in the South African Journal of Economics.

Riddell, W. C. and X. Song(2011). “The Impact of Education on Unemployment Incidence and Re-Employment Success: Evidence from the U.S. Labour Market.” IZA Discussion Paper. No. 5572.

Sargan, J. D.(1958). “The Estimation of Economic Relationships Using Instrumental Variables.” *Econometrica*. Vol. 26. No. 3, pp. 393-415.

World Bank(2013). World Development Indicators (ADI). Available online at http://data.worldbank.org/data-catalog/world-development-indicators. (Accessed on June 15, 2013)

| **논문투고일** : 2013년 07월 03일 |
| **논문심사일** : 2013년 08월 18일 |
| **게재확정일** : 2013년 08월 16일 |

ABSTRACT
Journal of Asia-Pacific Studies Vol. 20. No. 2 (2013)

Linguistic Diversity, Unemployment and Schooling:
Empirical Evidence from the East-Asian Countries

Alexandre Repkine
(Dept. of Economics, Konkuk Univ.)
Seung-Hoon Song
(Dept. of International Trade, Konkuk Univ.)

We explore both direct and indirect effects of linguistic diversity may produce on the labor market outcomes, in particular unemployment. For this purpose we use a measure of linguistic diversity that takes account of the languages' mutual intelligibility with a sample of seventeen East and South Asian countries to test the hypotheses of skill complementarity, ethnocentric behavior, social cohesion, and political aspirations. The latter two hypotheses correspond to the indirect effects of linguistic diversity on unemployment working through the educational system that encourages social integration of the minority language groups on the one hand, while increasing the risk of social conflict due to the ethnolinguistic differences. We find significant support for the hypotheses of ethnocentric behavior and political aspirations with the exception of primary school graduates for whom there is evidence of skill complementarities effect at work.

Key words: Unemployment, Linguistic Diversity, Education, Skills Complementarity, Ethnic Conflict