VYTAUTAS MAGNUS UNIVERSITY
M C M X X I I

ISSN 2335-2019 (Print), ISSN 2335-2027 (Online)
Darnioji daugiakalbystė | Sustainable Multilingualism | *27*/2025
https://doi.org/10.2478/sm-2025-0013

Milda Kiškė
Vilniaus kolegija / Higher Education Institution, Lithuania

LANGUAGE PROFICIENCY AND WAGE PREMIUMS: A REVIEW OF ANALYTICAL MODELS WITH REFERENCE TO LITHUANIA

Abstract. This paper reviews multiple cross-national empirical studies and proposes a comparative analytical framework for the economic value of multilingualism in the Lithuanian labour market. The study draws on the concepts of human capital theory, language economics and communication efficiency models. By synthesising the main findings on multilingual labour markets, the paper shows how wage differentials related to language skills can be analysed using different statistical methods. The study uses an illustrative simulated data set based on the PIAAC methodology. The data indicate that individuals with proficiency in the national language (host-country) skills can expect an average wage premium of 18.8%. The findings of the study suggest that to strengthen the workforce in Lithuania, it would be beneficial to integrate language policy with educational and employment systems. The statistical results presented in this paper are based on a simulated dataset modelled on the PIAAC framework and do not represent empirical findings from real Lithuanian labour force data. Additionally, the proposed analytical model provides a novel perspective and addresses a specific gap in existing research on the economics of multilingualism in Lithuania. The framework highlights policy levers such as transparent language requirements, targeted L2 support, and recognition of minority language capital—that can operationalise sustainable multilingualism, i.e., maintain linguistic diversity while improving equitable labour-market access in Lithuania.

Keywords: economic value; labour market; language skills; multilingualism; wage premium.

Introduction

Background

The economic value of multilingualism has attracted growing scholarly attention. Recent studies have shown a direct relationship between language skills, income, employment prospects, and career opportunities. This paper reviews the most cited and influential research studies in the field of language economics and multilingualism, serving as the foundation for the current research study.

The studies reveal that multilingual labour markets, such as those examined in Québec, Wales, Switzerland, Catalonia, Estonia, Latvia, Germany, Italy and Poland, create a positive environment in which foreign language proficiency offers a clear competitive advantage. Researchers have attempted to measure the wage premiums related to multilingualism, both for native speakers and immigrants. Together, these studies show that language skills often bring significant economic benefits (Chiswick & Miller, 1995; Di Paolo & Raymond, 2012; Albouy, 2008). However, the studies examined revealed that the economic benefits gained from language skills heavily depend on institutional, cultural and regional factors.

This gap presents an opportunity to review broader hypotheses of analytical models in language economics with reference to Lithuania. However, the findings of methodological approaches which aimed to determine the relationship between wage premiums and individual economic outcomes are difficult to systemise, because they often differ drastically due to the various analytical techniques and contexts used. The key challenges that the scholars faced were issues such as endogeneity, selection bias, and unobserved heterogeneity. Thus, to overcome the above-listed issues, researchers have employed tools such as Two-Stage Least Squares (TSLS), Propensity Score Matching (PSM), and fixed-effects panel models. For example, TSLS utilises special variables to help identify cause-and-effect relationships, while PSM compares individuals with similar characteristics but different language skills to examine how language affects outcomes.

Hence, keeping the above-mentioned issues in mind, this paper does not aim to produce new empirical data; instead, it seeks to provide

a conceptual and methodological synthesis of how language proficiency relates to wage premiums in the Lithuanian and broader European labour markets. By reviewing analytical models used in previous empirical studies, the paper establishes a framework that may later be applied to international and Lithuanian datasets. The goal is to identify theoretical approaches that can inform future empirical research in multilingual economic contexts. In addition, when determining economic outcomes, it is essential to emphasise the crucial role of institutions and cultural factors. For example, policies such as Bill 101 in Québec and the official recognition of Catalan in Spain have significantly influenced wage premiums and employment opportunities for speakers of minority languages (Shapiro & Stelcner, 1997; Di Paolo & Raymond, 2012; Rendon, 2007). The framework developed in this paper focuses on cross-regional comparisons. It provides practical guidance for policymakers and researchers seeking to enhance labour market inclusion and community cohesion in multilingual contexts.

The Lithuanian Sociolinguistic and Policy Context

Although a majority of the Lithuanian population knows at least one foreign language, nine out of ten people of working age state that they speak at least one foreign language (Eurostat, 2019); official systems, however, remain monolingual, recognising only Lithuanian. The Law on the State Language (Valstybinės kalbos įstatymas, 1995) formally establishes Lithuanian as the only official language in important sectors, excluding other languages from official recognition. According to this legislation, the Lithuanian language must be used throughout all contexts, from public institutions and official communication to educational settings. On the one hand, minority languages (such as Russian, Polish, or Belarusian) are allowed only in cultural spaces, not in work or state settings. Speakers of minority languages often lack official support in employment settings, which affects their access to wages and jobs (Seimas of the Republic of Lithuania, 1995–2023). On the other hand, substantial Russian-speaking communities remain, and the younger generation increasingly turns to English for socio-economic mobility. Despite this linguistic diversity, empirical assessments of how multilingualism affects economic

outcomes in Lithuania remain scarce, making it an ideal setting to test the broader hypotheses of language economics.

In contrast to Switzerland, where multilingual legal framework of several languages is officially recognised, or in Catalonia, where both languages Catalan and Spanish have equal status and economic value with a presumption that language proficiency in either can lead to better employment outcomes (Di Paolo & Raymond, 2012), Lithuania, by recognising only one official language, increases the economic importance of knowing Lithuanian.

While the Law on the Legal Status of Foreigners (further referred to as the Foreigners' Law) (Seimas of the Republic of Lithuania, 2004) does not explicitly mention language acquisition; its references to integration and naturalisation, aligned with EU frameworks, implicitly include linguistic and cultural adaptation. Following the Russian invasion of Ukraine, Lithuania implemented the "Language Learning as Part of Successful Social Integration" project (2020–2022) to support migrant integration, targeting third-country nationals and language teachers. The initiative, inspired by models from Austria, Germany, the Netherlands, and Sweden, aimed to enhance access to education and employment through structured Lithuanian language instruction (Lietuvos Respublikos socialinės apsaugos ir darbo ministerija, 2018). Furthermore, Lithuania's parliament has recently passed amendments to the State Language Law (Seimas of the Republic of Lithuania, 1995–2023) that will require foreigners and the companies employing them to serve clients in Lithuania in the national language starting in January 2026. The legislation aims to enhance communication and customer service in Lithuania while striking a balance between the needs of foreign employees and businesses. The strict language policy makes language more than just a communication tool; it is also a structural access mechanism to economic inclusion and upward mobility.

Between 2001 and 2020, Lithuania experienced the most significant relative decrease in demographics (−20%) and showed the highest increase in median age among other European Union countries during the period 2002–2022 (Eurostat, 2023). New migration inflows partially substitute for the decrease. Among newcomers to Lithuania, Belarusians made up the largest

group, with 15,675 arrivals in 2023, followed by Ukrainians (10,539) and Kyrgyz (3,227). Despite ongoing labour shortages in such sectors as cleaning, food service, manufacturing, healthcare, education, and construction, the Lithuanian government has recently tightened employment rules for foreign employees (European Labour Authority, 2023). Currently, only individuals with residence permits are allowed to work in the country. However, certain employees, including teachers, researchers, and citizens from economically developed countries, are exempt from this policy and have easier access to the Lithuanian job market.

Lithuania continues to face labour shortages in 59 different occupations, as noted in a recent report. However, the government is seeking to balance these shortages with stricter rules for foreign employment, aiming to maintain control over its workforce and ensure that citizens are served in the national language (European Labour Authority, 2023). As migrants are funnelled into low-skill or non-client-facing roles without language training, those with stronger Lithuanian skills command clear wage premiums and better career mobility.

Together, these dynamics position Lithuania as a compelling case where monolingual legal structures, migrant integration policies, and demographic trends converge to institutionalise the economic value of Lithuanian proficiency. Unlike multilingual societies where legal pluralism dilutes the dominance of a single language, Lithuania's unitary framework sharply accentuates the link between language skills and earnings potential, aligning with the wage premium trends illustrated in this paper's simulated model.

Objectives and Research Questions

The main *objective* of this study is to provide a conceptual and methodological synthesis of the analytical models used to assess the economic value of multilingualism. Specifically, the study seeks:

1. to review how various statistical models have been applied to estimate wage premiums associated with multilingualism;
2. to integrate these models into a unified framework that improves coherence and comparability;

3. to identify gaps in existing research and suggest methodological improvements for future studies.

To guide this analysis, the following *research questions* were formulated:

1. How have different analytical and statistical models been employed to measure the economic value of multilingualism and wage differentials across countries?
2. What methodological patterns and key variables emerge from existing studies on language proficiency and earnings?
3. To what extent can these models be adapted or applied to analyse Lithuania's multilingual labour market?
4. How can the synthesis of previous approaches contribute to a unified conceptual framework for future empirical research?

The paper is organised as follows. First, it defines and critically examines the concept of multilingualism, highlighting its evolving interpretations within the academic literature. Second, it examines the wage premiums earned by immigrants in host countries, with a focus on the impact of language skills on labour market integration. Third, attention shifts to native speakers in multilingual regions, examining how language competencies function as economic capital. Fourth, the paper presents the methodology, discussing in detail the statistical techniques employed to synthesise previous research. Finally, the paper concludes with key insights and policy implications for researchers and decision-makers, aiming to promote more inclusive multilingual labour markets.

Literature Review

Changing Concept of Multilingualism

The rise of digital communication tools and globalisation has significantly expanded our understanding of multilingualism over the past few decades (Hahm & Gazzola, 2022). Previously, the key function of

multilingualism served as an intercommunicational tool and cognitive asset to enhance intercultural awareness and facilitate communication (Gazzola & Mazzacani, 2019). However, due to the intensive interconnections in intensely globalised labour markets, the function of multilingualism has expanded beyond intercultural communication to promoting economic mobility, enhancing workforce adaptability, and informing language-related policy decisions.

Studies have shown a direct relationship between language skills, a form of human capital, and productivity and employability (Becker, 1993; Chiswick & Miller, 1995). Different communicative environments are more user-friendly to multilingual individuals, as their functional ability to manipulate more than one language in turn strengthens their position in the labour market. In addition to these functional benefits, a multilingual individual is also regarded as more successful in cognitive work-related environments, such as those that value flexibility, creativity, and problem-solving ability, qualities that are especially valued in knowledge-intensive sectors of the modern economy (Budría et al., 2019).

The economic benefits of multilingualism are especially well recognised in countries where legal frameworks and government programs determine language diversity. For example, a law called Bill 101 in Québec was designed to protect the status of the French language. As a result, individuals proficient in both French and another language tend to gain higher wage premiums compared to those who speak only one language (Shapiro & Stelcner, 1997; Albouy, 2008). Similarly, in Catalonia, the established language policy has helped individuals proficient in Catalan (and Spanish) gain a competitive advantage (Di Paolo & Raymond, 2012).

The concept of multilingualism has expanded over recent decades. The studies highlight the need to reconsider the analytical frameworks that embrace both economic returns and socio-cultural linguistic functions (Budría et al., 2019; Hahm & Gazzola, 2022).

In recent years, researchers have employed various theories, including Human Capital Theory, Language Economics Theory, and Communication Efficiency Models, to understand how knowing more than one language affects individuals in the labour market. Table 1 summarises the main theories that

scholars used to determine how multilingualism can create economic value in the workplace.

Table 1

Key Theoretical Approaches Explaining the Economic Value of Multilingualism in the Workplace

Theory	Key Concepts	Application to Multilingualism	Supporting Studies
Human Capital Theory	Skills as economic assets; productivity enhancement.	Language proficiency enhances productivity, adaptability, and labour market integration.	Becker (1993); Chiswick & Miller (2003).
Language Economics Theory	Direct and Indirect Value of Language Skills.	Provides immediate returns through employability and wage premiums; enhances other human capital forms.	Budría et al. (2019); Zorlu & Hartog (2018).
Communication Efficiency Models	Reduced communication barriers, increased productivity.	Language proficiency improves information exchange and market access.	Stöhr (2015); Isphording (2015).

Human Capital Theory views language proficiency as a valuable skill, demonstrating that an individual with good language skills is more productive and inclusive, especially in multilingual work environments. This suggests that language proficiency influences the wage premium and better employment opportunities (Becker, 1964; Chiswick & Miller, 1995). Language Economics Theory categorises the economic value of language skills into two categories: direct benefits and indirect benefits gained from multilingualism in the labour market (Budría et al., 2019; Zorlu & Hartog, 2018). The tangible benefits refer to higher personal incomes and more fluent participation in job opportunities. More transparent communication and easier knowledge transfer increase work productivity in the workplace and are referred to as the intangible benefits gained from multilingualism in the job market (Budría et al., 2019; Zorlu & Hartog, 2018). Thus, it is suggested that language skills increase in value when combined with education and job-specific skills. Furthermore, Communication Efficiency Models emphasise that speaking multiple languages reduces misunderstandings and enhances collaboration, thereby increasing overall efficiency in multilingual workplaces (Stöhr, 2015; Isphording, 2015).

Wage Premiums Earned by Immigrants in Host Countries

Wage premiums earned by immigrants in host countries are one of the clearest indicators of the economic value of multilingualism, resulting in higher incomes, better labour market mobility, and smoother integration into the host country's social and economic environment (Chiswick & Miller, 1995; Budría et al., 2019). The findings from labour markets in regions such as North America, Europe, and Australia suggest the importance of investing in human capital to improve productivity and help immigrant employees succeed in the labour market, as well as to prove the relationship between language acquisition and labour market outcomes (Chiswick & Miller, 1995). Studies show that highly skilled immigrants with strong command of the dominant language earn more and transition jobs more readily.

The height of wage premiums for immigrants from language skills varies significantly depending on the specific context of the host country. For example, cases in Canada and the United States proved that proficiency in the host language strongly correlates with increased earnings (Albouy, 2008; Shapiro & Stelcner, 1997; Chiswick & Miller, 1995). Strong language skills in either English or French among immigrants in Canada have been associated with higher earnings compared to those with limited language skills (Albouy, 2008; Shapiro & Stelcner, 1997). Similarly, in the United States, recent immigrants who were proficient in English and had not yet established themselves in the labour market earned more than those with no skills in the dominant language (Chiswick & Miller, 1995). While the current research focuses on economic factors related to language acquisition, studies in the field are showing a growing recognition of the broader social benefits of language proficiency among immigrants.

Wage Premiums Earned from Language Skills for Natives in Multilingual Markets

The second strand of existing literature examines the benefits that natives derive from foreign language proficiency in bilingual or multilingual labour markets, particularly in environments where individuals are proficient in

both regional and national languages (Shapiro & Stelcner, 1997; Grin & Sfreddo, 1998; Di Paolo & Raymond, 2012). Cases from countries such as Canada, Wales, Switzerland, Catalonia in Spain, as well as other European countries, have sparked the interest of researchers in the field.

To estimate wage differentials related to language skills, researchers employed several quantitative methods. Shapiro and Stelcner (1997) analysed the impact of bilingualism on income in Québec using a specialised statistical model that helped to address problems when the data were biased or not fully representative. They employed a two-step method that included a so-called probit model to address the issues arising from the selection of data, thereby allowing them to analyse wage differences between language groups over 20 years using census data from 1971, 1981, and 1991. The study results revealed that bilingual Francophone individuals gained significant economic benefits, while monolingual Anglophones, particularly those working part-time, were at a relative disadvantage. The study also highlights the role of Québec's 1977 Bill 101, which had a significant impact on the use of the French language, thus creating more possibilities for French-speaking employees in the job market.

Similarly, to determine the wage differences between French-speaking and English-speaking employees, Albouy (2008) combined three methods: Mincer-type wage regressions, Oaxaca-Blinder decomposition, and quantile regressions. First, basic wage regressions were conducted to examine how factors such as language affect income. Second, a method called Oaxaca-Blinder decomposition helped determine what part of the wage gap is due to fundamental differences (such as education or experience) and what part cannot be easily explained (possibly due to discrimination or other hidden reasons). Third, quantile regressions were used to examine whether the wage gap varies depending on whether someone earns a low, middle, or high income. Using census data from 1970 to 2000, the study showed that the pay gap between French and English speakers in Québec decreased by about 25 percentage points. In the rest of Canada, the gap narrowed by only about 10 points. The study findings suggest that this change in Québec was not because French speakers started earning a lot more—it was mainly because English speakers in Québec started earning less. Albouy (2008) explained that

this occurred due to changes in Québec's economy and a key language law, Bill 101, introduced in 1977, which gave the French language a stronger role in society. His study reveals that language-related income disparities often depend on local government policies and broader social structures.

In their study, Grin and Sfreddo (1998) examined the wage premium that individuals in Switzerland earn based on the language they speak. Using a statistical analysis called OLS regression, Grin and Sfreddo (1988) examined the impact of education, job experience, and language skills on wage premiums. Focusing on Italian speakers, the study compared their wages to those of German and French speakers, aiming to test whether multilingual employees (especially those with Italian plus another official language) received wage premiums. Moreover, to make their results more accurate, they examined residents living in different parts of Switzerland and also separated the data between all residents and Swiss citizens only. Their results showed that Italian speakers often earn less, especially if they live in areas where Italian is not the main language. Hence, the results suggest that a minority language may still face disadvantages even in a country with multiple official languages. However, the study acknowledges that its data has some limitations regarding other important factors that might affect the results.

Using census data from 1991 and 1996, Rendon (2007) found proper evidence regarding the influence of language skills on employment and income in the Catalan region. The study findings showed that employability possibilities increased by 3–5 percentage points for people who could read and speak fluently in Catalan, while those who could also write in Catalan saw a 2–6 percentage point advantage. Later, Di Paolo and Raymond (2012) employed the Two-Stage Least Squares (TSLS) method to exclude any biases in the results where language skills and wages might influence each other. The study findings showed that skills in the Catalan language increased monthly earnings by approximately 7.5% under the standard regression, and the percentage grew to 18% after controlling for bias and possible errors. Hence, the study findings suggest that fluency in the Catalan language offers clear economic benefits in the region.

Table 2 compares the wage premiums earned from language skills in different regions. The table includes information such as language laws in

the area, the importance and prevalence of the language in the region, the methodological approaches employed, and the limitations of the research. This comprehensive summary contributes to a better acknowledgement of how the economic returns to language skills vary depending on the regional context and the sociolinguistic status of the language.

Table 2

The Comparative Analysis of Wage Premiums Earned from Language Skills Across Various Regions

Region	Wage Premium Range (%)	Statistical Models Applied	Key Findings	Limitations
Québec (Canada)	25 (Francophones)	OLS, Probit, TSLS	Bilingualism benefits Francophones; Anglophones face disadvantages. Bill 101 enhanced the economic status of Francophones.	Limited generalizability beyond Québec; focus on Francophones and Anglophones.
Wales (UK)	8–10	Econometric Models	Bilingual workers earn approximately 8–10% more than monolinguals.	Small sample size; limited sectoral representation.
Switzerland	-5 (Italians)	OLS Regression	Italian speakers face economic disadvantages compared to German and French speakers.	Inconsistent statistical significance across models.
Catalonia (Spain)	7.5–18	TSLS, OLS	Catalan proficiency increases employment probability by 3–6 percentage points. Accounting for endogeneity increases wage premiums from 7.5% to 18%.	Regional specificity; limited generalizability.
Estonia & Latvia	10	Econometric Analysis	Local language proficiency increases income, particularly in the public sector.	Limited applicability to private sector economic returns.

Region	Wage Premium Range (%)	Statistical Models Applied	Key Findings	Limitations
Germany, Italy, Spain	13 (Germany)	Probit Models	English proficiency enhances employment prospects; French proficiency is negligible.	Limited discussion on societal and cultural factors.
Poland & Abroad	6–22	Survey-Based Analysis	Foreign language proficiency offers a 6% wage premium in Poland and a 22% premium abroad. Higher premiums for Spanish, Italian, and French than for English.	Lack of a detailed breakdown of contributing factors.

The table illustrates how wage premiums from language proficiency vary according to different factors. For example, after the establishment of Bill 101, in Québec, people with skills in both French and English tend to earn more than those who speak only one language (English speakers) (Shapiro & Stelcner, 1997; Albouy, 2008).

As has been shown, in Catalonia, knowledge of Catalan enhances both job opportunities and earnings, especially after controlling for any biases where language skills and wages might influence each other (Di Paolo & Raymond, 2012). In comparison, the study findings from Estonia and Latvia show that wage premiums are only evident in the public sector, thus suggesting that economic rewards from language skills are highly dependent on government policies and formal language rules (Toomet, 2011; Ridala, 2020).

Furthermore, the table confirms the notion that the English language remains the dominant language in the international business world (Gazzola & Mazzacani, 2019; Liwiński, 2018). The study findings from Germany, Poland, and other parts of Europe support this fact, showing that proficiency in English brings obvious financial advantages (Gazzola & Mazzacani, 2019; Liwiński, 2018). On the other hand, the research findings suggest that proficiency in minority languages, such as Italian in Switzerland, may earn less outside their

home language regions (Grin & Sfreddo, 1998; Cattaneo & Winkelmann, 2005).

Taken together, these comparisons confirm the interdependability between how local language policies and legal frameworks influence the economic value of language skills in the labour market. Thus, there is a need to standardise research methods, which would create a better environment for comparing and interpreting findings across different countries and regional contexts.

Although numerous international studies have examined language proficiency and wage premiums, research in the Lithuanian context is minimal. To the best of the author's knowledge, there are no comprehensive Lithuanian-language studies that directly analyse the economic value of multilingualism. Existing publications tend to focus on education or migration rather than quantitative labour-market outcomes. Similar gaps exist in other Baltic states. Therefore, this review aims to consolidate existing analytical approaches and provide a foundation for future empirical investigations in Lithuania and the Baltic region.

To address these limitations, this study introduces a unified framework that integrates diverse statistical models, standardises data collection practices, and incorporates the economic, social, and cultural dimensions of multilingualism. This approach offers a more comprehensive understanding of how language proficiency affects wage outcomes and employment opportunities across diverse regional contexts.

Methodology

This methodological framework presents a structured review of the statistical models used to estimate the economic value of multilingualism. It is important noting that these models are presented illustratively to show how prior studies addressed endogeneity, selection bias, and unobserved heterogeneity. Instead, the aim here is to explain how past studies have applied these tools to overcome the common research problems, such as biased samples (endogeneity), hidden variables, or unclear cause-and-effect relationships (heterogeneity).

Data Collection and Standardization

This review integrates data from various sources, such as national labour market surveys, longitudinal studies, structured questionnaires, and data from legal documents. By combining these different types of data, it offers a comprehensive analysis of how wage premiums from multilingualism vary across regions.

Table 3 provides a structured summary that introduces the type of data used (for example, job surveys, long-term studies, questionnaires) and statistical methods employed to analyse that data (such as regression models or matching techniques) to determine wage premiums from language proficiency. To better compare research results in the field conducted across different regions, we emphasised the aim and limitations of the studies, as well as the benefits gained from the language proficiency. The existing literature in the focus field employs various types of data and statistical approaches to determine the economic value of language skills. Findings from several studies are based on national labour market data, allowing for the analysis of large-scale cross-sectional data (Chiswick & Miller, 1995; Albouy, 2008; Gazzola & Mazzacani, 2019). Techniques such as Propensity Score Matching (PSM), Two-Stage Least Squares (TSLS) and fixed or random effects models were commonly employed to determine the differences in wage premiums depending on language knowledge.

The studies by Aldashev et al. (2009) and Ridala (2020) utilised panel data to examine the dynamics of how language skills influence career possibilities over a longer time period.

Structured questionnaires applied by Budría et al. (2019) and Liwiński (2018) focused on qualitative language use. The authors applied PSM and probit models and compared wage premiums between groups with differing language skills.

Table 3

Summarised Data Sources and Statistical Methods Applied by the Reviewed Studies to Analyze Wage Premiums from Language Skills

Data Source	Variables Collected	Measurement Techniques	Statistical Models Applied	Purpose	Benefits	Limitations	Supporting Studies
National Labour Market Surveys	Language Proficiency, Earnings, Employment Status, Education, Demographics, Occupation.	Structured Surveys, Government Reports (e.g., OECD, Eurostat, ILO).	PSM, TSLS, Fixed/Random Effects.	Cross-sectional analysis of language-related wage premiums.	Large sample size, statistical precision.	Limited data on individual progression over time.	Chiswick & Miller (1995); Gazzola & Mazzacani (2019); Albouy (2008).
Longitudinal Studies	Language Skills (Over Time), Career Progression, Earnings Growth, Occupational Mobility.	Repeated Surveys, Panel Data Sets (e.g., Germany, Eurostat Panels).	Panel Data Analysis, TSLS, Fixed-Effects Models.	Dynamic analysis of language skills' impact over time.	Tracks individual progression, reduces bias.	Limited availability of longitudinal data.	Aldashev et al. (2009); Ridala (2020); Di Paolo & Raymond (2012).
Structured Questionnaires	Language Proficiency (Self-Assessment), Social Capital, Cultural Integration, Perceived Language Value.	Self-Reported Surveys, Interviews, Focus Groups.	PSM, Probit Models.	Comparison of language-related premiums across groups.	Controls for observable characteristics.	Susceptible to self-reporting bias.	Budría et al. (2019); Liwiński (2018); Toomet (2011).
Institutional Data	Historical Language Policies, Regional Legal Frameworks.	Official Documents, Legal Reports.	Fixed/Random Effects Models, TSLS.	Analyzing institutional impacts on wage premiums.	Captures legal and policy effects.	Limited comparability across regions.	Shapiro & Stelcner (1997); Grin & Sfreddo (1998); Cattaneo & Winkelmann (2005).

Several research papers aimed to establish a better understanding of how language policies and legal rules across different regions influence economic outcomes over time. Shapiro & Stelcner (1997), Grin & Sfreddo (1998), and Cattaneo & Winkelmann (2005) applied the Two-Stage Least Squares (TSLS) method to explore whether an individual earns higher premiums due to language proficiency or whether other hidden factors influence the results.

The various methods and data sources used in the existing literature of the focus field make the findings summarised in the table more credible and reliable across multiple multilingual contexts.

Statistical Models and Techniques

This section discusses how different statistical models were used to analyse the relationship between language skills and wage differentials. The formulae presented below serve to solely illustrate the types of methods that other researchers have used in similar studies, as well as the references to existing methodologies applied by previous researchers.

Two-stage Least Squares (TSLS)

Researchers use the TSLS method in the literature reviewed when they suspect that undefined factors, such as motivation or socio-economic background, could distort the results when trying to determine how strongly language skills influence higher income. By using instrumental variables (IVs), this method helps filter out factors related to language skills but not directly related to income. The TSLS model comprises two steps. In the first step, the model predicts language proficiency using the instrumental variables. In the second step, this prediction is used to estimate the actual impact of language skills on income, without the bias from the hidden factors. Illustrative Formulas (For Reference Only):

1. Instrumental Variable Regression:

$$Li = \alpha 0 + \alpha 1 Zi + \alpha 2 Xi + ui$$

Where:

Li = Predicted Language Proficiency.

Zi = Instrumental Variable (e.g., Age at Immigration, Parental Language).

Xi = Control Variables (e.g., Education, Gender, Occupation).

ui= Error Term.

2. Second Stage (Outcome Regression):

$$Yi = \beta 0 + \beta 1 Li\hat{} + \beta 2 Xi + \epsilon i$$

Where:

Yi = Economic Outcome (e.g., Earnings).

$Li\hat{}$ = Language proficiency as predicted by the first regression

ϵi = Error Term.

In their research, Di Paolo & Raymond (2012) used TSLS to determine wage premiums regarding Catalan proficiency. Researchers noticed that language skills might be connected to other hidden factors (like motivation or family background), which could distort the results when trying to see if speaking a language leads to higher income. This is referred to as a problem of endogeneity.

To overcome this problem, they employed a special method called Two-Stage Least Squares (TSLS). Instead of directly using language skills, they first predicted language ability based on something not directly related to income, in this case, how much the person was exposed to the language in the past (this is called an instrumental variable). Using ordinary least squares (OLS) regression, they found that speaking the local language led to a 7.5% higher income. However, when they used the TSLS method (which better isolates cause and effect), the estimate jumped to 18%. This means that after correcting for hidden influences, the effect of language skills on wages turned out to be even bigger.

Similarly, Albouy (2008) employed the same method to investigate income differences between French and English speakers in Québec, aiming to

understand the impact of language skills alone on earnings, without other factors interfering.

Propensity Score Matching (PSM)

The PSM method compares individuals with similar characteristics but differing language skills.

Illustrative Formulas (For Reference Only):

1. Estimating Propensity Scores:

$$P(Ti) = Pr(Ti = 1 \mid Xi)$$

Where:

Ti = Treatment Variable (1 if proficient, zero if not).

Xi = Vector of Control Variables (e.g., Education, Age, Occupation).

2. Matching Process:

Nearest Neighbour Matching: Compare each treated individual with the most similar untreated individual.

Kernel Matching: Use a weighted average of all individuals based on their propensity scores.

3. Treatment Effect Calculation:

$$ATE = E[Yi(1) - Yi(0)]$$

Where:

$Yi(1)$ = Outcome for treated individuals.

$Yi(0)]$ = Outcome for untreated individuals.

Chiswick and Miller (1995) and Budría et al. (2019) both used a method called Propensity Score Matching (PSM) to make fair comparisons between people with different language abilities. Chiswick and Miller (1995) looked at immigrants in the U.S. and compared how much bilingual immigrants earn versus monolingual ones, making sure the two groups were similar in other ways (like age, education, etc.). Budría et al. (2019) conducted a similar study in Europe, comparing wages across individuals from different language groups while matching them with similar backgrounds to ensure a fair comparison. It

helped to isolate the effect of language skills by making sure the people being compared are alike in all other important ways. Therefore, if there is a difference in wages, it is more likely due to language ability, rather than something else.

Panel Data Analysis

Panel data analysis is employed to capture dynamic effects over time. Illustrative Formulas (For Reference Only):

1. Fixed-Effects Model:

$$Yit = \alpha i + \beta Lit + \gamma Xit + uit$$

Where:

Yit= Outcome Variable (e.g., Earnings) for individual iii at time ttt.

Lit = Language Proficiency.

Xit= Time-variant Control Variables (e.g., Employment Type, Education).

αi = Individual-Specific Effect (constant over time).

UIT = Error Term.

2. Random-Effects Model:

$$Yit = \alpha + \beta Lit + \gamma Xit + \mu i + uit$$

Where:

Yit =Earnings for individual at time i at time t

Lit Language proficiency

Aldashev et al. (2009) investigated the economic returns of language proficiency among foreigners in West Germany. They found that language skills have a positive influence on employment probabilities and occupational choices, particularly in white-collar professions. However, the direct effect of language proficiency on earnings diminished when controlling for occupational selection and economic sector characteristics. Ridala (2020) applied panel data models to assess the impact of multilingualism on labour market outcomes in Estonia and Latvia.

Fixed and Random Effects Models

To control for unobserved heterogeneity across individuals and regions. Illustrative Formulas (For Reference Only):

1. Fixed-Effects Model:

$$Yi=\alpha i+\beta Li+\gamma Xi+\epsilon i$$

2. Random-Effects Model:

$$Yi = \alpha + \beta Li + \gamma Xi + \mu i + \epsilon i$$

Grin & Sfreddo (1998) and Cattaneo & Winkelmann (2005) applied these models to analyze wage differentials in Switzerland.

The application of these statistical models provides a comprehensive framework for assessing the economic value of multilingualism. By examining how these methods have been used in various studies, this review identifies best practices and methodological gaps that future research should address.

Comparative Analytical Framework

To understand how multilingualism affects economic, social, and cultural outcomes, researchers employ a structured approach that incorporates various statistical methods. Comparing outcomes across countries or regions becomes easier when this structure highlights how factors such as government policy, cultural context or economic conditions influence the income associated with language skills. Many of these studies employ methods such as TSLS, PSM, panel data analysis, and fixed or random effects — not only to avoid bias but also to enhance the accuracy of the results.

Empirical Methods

Based on data from Lithuania, particularly from employed persons aged 25 to 65, this study demonstrates how the analytical framework operates. The information comes from the OECD's PIAAC survey (OECD, 2016), which collects details about people's work, skills and personal background across countries.

To create the model, this Lithuanian subset was taken from the larger international dataset. Monthly income (G_Q07a_REC), the most important variable in the analysis, was taken from a survey in which income was expressed in ranges. To estimate a single number for each person, the researchers used the mean value of each range. Then, to make the numbers easier to work with and compare, especially because some incomes are much higher than others, they transformed the income values using a mathematical tool called the natural logarithm. This smoothes out the differences between low and high incomes, making the patterns in the data easier to identify and analyse.

To determine language proficiency, three specific questions were asked in the survey: the language in which the test was taken (L_TestLang), the language spoken at home (B_Q01a), and how well respondents reported they could speak the test language (L_SpeakLangTest). Respondents were considered fluent in the national language if they completed the test in Lithuanian and stated that they spoke Lithuanian at home. In addition to language skills, other background factors such as educational level based on ISCED categories (B_Q06a), age (AGE_R) and gender (GENDER_R) were also a part of the analysis.

The current paper used version 2.3 of the PIAAC dataset (published in 2023) for this analysis. However, the background questionnaire containing the survey questions was last updated in 2012. Lithuania participated in Round 1 of PIAAC, thus there is no newer version of the questionnaire for the Lithuanian data. Hence, all survey items and variable references used in this study are based on the 2012 background questionnaire, which is the standard for all countries in Round 1, including Lithuania.

All data simulations and statistical analyses in this study were conducted using R software (version 4.3.3). Core functions employed include lm() for Ordinary Least Squares regression and base R data generation functions for simulation. This approach ensures transparency and reproducibility in the modelling of wage differentials related to language proficiency. In addition to language skills, the model also included other important factors that could influence salary, such as age, gender and education level, so that the impact of language skills could be

measured more clearly. The model specification is:

$$\log(\text{wage}) = \beta_0 + \beta_1(\text{Language\_Proficiency}) + \beta_2(\text{Education}) + \beta_3(\text{Gender}) + \beta_4(\text{Age}) + \varepsilon$$

The model, therefore, attempts to measure the contribution of language knowledge to income after accounting for the effects of education, gender, and age. The illustration is consistent with prior evidence that host-language proficiency is associated with higher wages.

It is worth noting that while the OECD's Programme for the International Assessment of Adult Competencies (PIAAC) provides rich microdata across countries, key variables for Lithuania, including wage measures and language background indicators, are missing from both the Public Use File (PUF) and complete datasets (OECD, 2016). As a result, this paper employs a simulated dataset modelled on the PIAAC framework to illustrate the analytical methodology.

This approach provides a transparent and replicable foundation for future empirical research once the complete Lithuanian data become available. Simulation enables the exploration of theoretical relationships even in the absence of complete national microdata, offering a valuable interim solution for hypothesis testing and model demonstration.

Variables and Simulated Results

The primary objective of this study is to investigate the relationship between language proficiency and wage outcomes in multilingual labour markets. To support this objective, a small-scale empirical illustration was conducted using a simulated dataset modelled on the structure and coding of the OECD PIAAC dataset for Lithuania (OECD, 2016). While this simulation does not allow for empirical validation, it demonstrates how the proposed analytical framework can be operationalised using available international datasets. The analysis is based solely on simulated data reflecting the structure of the OECD PIAAC dataset for Lithuania.

A simplified regression model was simulated using realistic, PIAAC-inspired data to explore how language proficiency could affect wages in Lithuania. It is important to note that the calculations used in these sections are a simulation, not real-world data. It is based on the PIAAC variable structure but with randomly generated values. The goal is to illustrate how the model might work if we had actual Lithuanian microdata. The simulated regression was employed as described in this section below. First, we generated 500 hypothetical individuals aged between 25 and 65. Second, we assigned population-tested, realistic-looking values for income, education, age, gender, and language proficiency. The variables assigned to each person tested were as follows:

- Wages (G_Q07a_REC): Monthly gross earnings were simulated in brackets and recoded using midpoint values to approximate a continuous wage variable.
- Monthly income in brackets → converted to midpoints → then converted to log(wage). The natural logarithm of these values was used as the dependent variable to allow interpretation in percentage terms.
- Language Proficiency (L_SpeakLangTest, L_TestLang, B_Q01a): defined as "proficient" if a person took the test in Lithuania or spoke Lithuanian at home. This proxy represents host-country language proficiency, the main independent variable.
- Education (B_Q06a): simulated educational attainment was categorized into Low, Medium, and High using ISCED classifications and given its strong association with labour market outcomes.
- Age (AGE_R): simulated as a continuous control variable to account for experience and lifecycle wage effects.
- Gender (GENDER_R): simulated as a binary control variable (Male/Female) to reflect potential gender-based wage differentials.
- Employment Status (B_Q04a): all simulated individuals were considered employed, consistent with the real PIAAC filtering applied in Section 4.1.

Table 4

Variable Definitions and Roles in the Simulated Regression Model

Variable Name	PIAAC Code / Source	Description	Role
Wage	G_Q07a_REC	Monthly gross earnings (simulated using midpoints from income brackets)	Dependent variable
Language Proficiency	L_SpeakLangTest, L_TestLang, B_Q01a	Proficiency in Lithuanian based on test language and home use	Independent variable
Education (Low/Med/High)	B_Q06a	Simulated education levels using ISCED categories	Control variable
Age	AGE_R	Continuous age variable (25–65)	Control variable
Gender (Male/Female)	GENDER_R	Binary gender variable	Control variable
Employment Status	B_Q04a	All participants were considered employed	Sample filter only

Third, we converted income to log form so that effects are easier to interpret and be less distorted by very high wages. Fourth, we applied a basic Ordinary Least Squares (OLS) regression using the following formula:

$$\log(\text{Wage}) = \beta_0 + \beta_1(\text{Language Proficiency}) + \beta_2(\text{Education}) + \beta_3(\text{Age}) + \beta_4(\text{Gender}) + \varepsilon$$

This model allows to interpret each coefficient as an approximate wage premium.

Table 5

Results of the Simulated Ordinary Least Squares (OLS) Regression Model, Estimating the Relationship Between Language Proficiency and Wages

Variable	Coefficient (β)	Standard Error	p-value	Interpretation
Language Proficiency	0.188	(0.025)	< 0.001	~18.8% higher wages for individuals proficient in Lithuanian

Variable	Coefficient (β)	Standard Error	p-value	Interpretation
Education (Medium)	0.158	(0.020)	< 0.001	~15.8% higher wages than those with low education
Education (High)	0.326	(0.027)	< 0.001	~32.6% higher wages than those with low education
Gender (Male)	0.104	(0.018)	< 0.001	Males earn ~10.4% more than females
Age	Not significant	—	> 0.05	No measurable wage effect in the simulated 25–65 age range

These illustrative results support the findings from previous empirical studies (e.g., Chiswick & Miller, 1995; Di Paolo & Raymond, 2012), which report significant wage premiums associated with proficiency in the host country's language. Although the current model is based on simulated data, it provides a transparent and replicable pathway for future empirical research using actual PIAAC microdata for Lithuania.

The results presented in this study are entirely based on a constructed simulation modelled on PIAAC-style variables and distributions, not real-world data. While this simulation provides valuable insights into possible wage premiums for multilingual individuals in Lithuania, it does not constitute empirical validation. This model serves for an illustrative purpose, demonstrating how economic returns to language skills could be analysed if representative microdata were available. Future research should empirically validate these simulated findings using large-scale households or labour force surveys that include actual language proficiency and employment data. Access to nationally representative datasets, such as Lithuanian LFS or PIAAC microdata, would be essential for determining causal inference. These findings, although illustrative, can inform policymakers and educators on how language proficiency may impact wage equity and labour integration in multilingual contexts.

Discussion

The results of this literature review examined the appropriateness, effectiveness, and limitations of the statistical methods used in the scholarly literature of the focus field, which were used to measure the economic value gained from language skills. By synthesising research findings from Québec, Wales, Switzerland, Catalonia, Estonia, Latvia, Germany, Italy and Poland cases, it becomes evident that benefits of multilingualism highly depend on regional, cultural and institutional contexts. This section discusses how different statistical models have been applied to examine these differences, focusing on their applicability, strengths, and limitations.

Regional Comparisons and Statistical Models

After examining the existing literature in the focus field, it becomes evident that there is no single statistical method to determine wage premiums associated with language skills. However, while the most widely used methods, such as Two-Stage Least Squares (TSLS), Propensity Score Matching (PSM), Panel Data Analysis and Fixed/Random Effects Models, aim to achieve the same goal, each comes with its own purpose. Moreover, it is worth noting that each model helps correct or control for specific types of bias, which in turn increases accuracy, depending on the type of data source.

Researchers have employed Two-Stage Least Squares (TSLS) to address a problem called endogeneity, which occurs when the independent variable, such as language proficiency, is influenced by other unobserved factors (e.g., motivation, intelligence, or socio-economic status) that also affect the dependent variable (wages). In the case of the research analysed, TSLS works in two steps: predict language proficiency using the IVs. The second step is to use that prediction in the wage regression to get a less biased estimate. For example, Di Paolo & Raymond (2012) used OLS (ordinary least squares) first and found that in Catalonia, the OLS premium for Catalan proficiency (~7.5%) rose to ~18% under TSLS, consistent with endogeneity in language–earnings estimate. This suggests the real effect of Catalan proficiency on wages was underestimated when endogeneity was not addressed. Albouy (2008) also

used TSLS to study wage gaps in Québec. Their analysis showed that where people live (region) and language laws (e.g., Québec's Bill 101) have a significant impact on wage outcomes. This supports the idea that language-related economic benefits depend not just on language skills, but also on regional policy contexts. However, TSLS works only if the chosen instruments (IVs) are valid. A valid instrument must be strongly related to language proficiency and have no direct effect on wages (only through language). In multilingual societies, it is hard to find such clean instruments. For example, parental language or age at immigration might still be indirectly related to income. So, TSLS can produce biased results if the instrument is weak or invalid.

In contrast, Propensity Score Matching (PSM) allows for the comparison of populations that are significantly different in language knowledge but similar in ways that affect the outcome (e.g., education level, age, gender). It has already been successfully demonstrated that possessing more language skills often leads to higher wages, especially in regions where these skills are in demand. However, PSM requires large datasets and cannot account for hidden differences between individuals, which limits its usefulness when data are limited.

Chiswick and Miller (1995) and Budría et al. (2019) used PSM in their studies to compare wages between groups that differ in their language skills but are otherwise similar and found that knowledge of more than one language often leads to higher wages, especially in regions where multilingualism is required or valued (e.g. in international cities or border regions). However, to apply PSM, there must first be an adequate population size to find suitable matches for all characteristics (such as age, education, and gender). Second, unobserved heterogeneity refers to important differences that are not captured by the data (such as motivation, personality or informal networks). Third, if these unobserved factors influence wages, PSM cannot adjust for them, and the result may still be biased. Therefore, PSM is not ideal when applied to small or incomplete data sets.

Panel data analysis is particularly useful for understanding the long-term effects of multilingualism on income. This method has been used in places such as Germany, Estonia, and Latvia. Aldashev et al. (2009) applied this

method to demonstrate how the wage benefits of language proficiency evolved over time in Germany, while Ridala (2020) employed the same method in Estonia and Latvia to investigate how income changes in relation to language proficiency. Together, these studies show how panel data can reveal a dynamic or trajectory. Unlike cross-sectional studies (which take a single snapshot in time), panel data helps to track how the same person's income changes over time. This reduces bias because population compared may differ in many ways. Instead, using this method, the same population is observed over a long-time period. However, the main challenge is that such long-term data is not always available, which makes it hard to apply the results to wider populations.

Fixed and random effects models refer to the identification of differences between individuals or regions that cannot be measured directly (unobserved heterogeneity) when measuring how language skills affect wages. These models allow to make more reliable comparisons between people or places by reducing hidden biases. Grin & Sfreddo (1998) and Cattaneo & Winkelmann (2005) studies using Switzerland, a multilingual country with German, French and Italian speakers, as a case for their study, applied these methods to measure the wage gap influenced by language skills, with. Even in a country that officially supports multiple languages, the study found that linguistic minorities (e.g. Italian speakers) often earn less than majority language speakers. The use of fixed/random effects models helped to reveal consistent wage differences that could not be explained by other obvious factors. However, assumptions about the data is often a decisive factor whether to use fixed or random effects depends on the. For example, if individual characteristics (such as region or native language) are related to the independent variables (such as education or language skills), fixed effects are better, or if they are not related, random effects may be more efficient. This choice may affect the results, and the right decision may change depending on the country or context.

Institutional and Cultural Factors

The findings also emphasise that language proficiency alone does not determine its value. Institutions or official data (such as government laws and

language policies) and culture (attitudes toward languages or minorities) are decisive in determining whether and how language skills influence higher earnings. Two contrasting case studies by Shapiro & Stelcner (1997) and Albouy (2008) offer a positive example in Québec, illustrating how a strong language law (Bill 101) promoted the use of French in business and public life. This strengthened the positions of the French and improved the economic outcomes for francophones. The results demonstrate how a supportive institutional environment can enhance the return on investment in language skills—the opposite case is provided in the research by Grin & Sfreddo (1998). Italian speakers in Switzerland, although residing in an officially multilingual country, tend to earn less, especially when living in areas where Italian is not the dominant language. This suggests that official recognition is insufficient if the language is not strongly supported by policy or culturally valued.

Rendon (2007) and Di Paolo & Raymond (2012) in their research studies demonstrate that institutions such as the government and the education system support regional languages (like Catalan), which improves the economic opportunities of the population. They found that people who speak Catalan well are more employable in the labour market and earn higher wages. This shows that language proficiency in a regionally important language has measurable benefits for individuals. The researchers applied Two-Stage Least Squares (TSLS) to isolate hidden influences, making the wage effect of Catalan proficiency more trustworthy.

Similarly, Gazzola & Mazzacani (2019) found that English proficiency enhances employment opportunities in Germany, Italy, and Spain, particularly for men. However, the economic returns associated with other foreign languages, such as French and Spanish, are less pronounced, suggesting that the economic value of language skills is context-dependent and shaped by regional demand.

Integration of Statistical Models into the Unified Framework

The integration of statistical models into the unified framework enhances coherence and comparability across studies. TSLS, PSM, Panel Data Analysis, and Fixed/Random Effects Models are complementary rather than

mutually exclusive. For example, TSLS addresses endogeneity concerns, while PSM controls for selection bias, and Panel Data Analysis captures dynamic changes over time.

Moreover, the use of Fixed and Random Effects Models allows for cross-regional comparisons, accounting for unobserved heterogeneity that would otherwise bias estimates. The findings suggest that the application of these models should be tailored to specific contexts, considering the availability of data, the nature of language-related policies, and the socio-economic characteristics of the populations studied.

Limitations and Recommendations for Future Research

The extensive literature review revealed that several limitations remain, despite the methodological improvements made by various studies. The reliance on cross-sectional data in many studies limits the ability to track individual progress over time, particularly when assessing long-term economic outcomes associated with language proficiency. Although TSLS is used to solve the problem of endogeneity and provides a more accurate picture of the relationship, it is not always available for use because a valid instrumental variable is not always available.

To understand how language skills affect employees‘ economic progress over time, future studies should focus on expanding longitudinal datasets, thereby revealing cause-and-effect patterns, such as whether learning a language at a particular moment improves income over a longer period. Additionally, using standardised, comparable methods across studies or regions could facilitate more straightforward comparison of findings, data combination, and reliable conclusions. Involving qualitative factors such as attitudes, experiences, and perceptions is often excluded from statistical research. Furthermore, mixed-method approaches, combining both qualitative and quantitative methods, would help gain a more detailed understanding of how language ability affects income, employment, and social mobility. At the same time, recognising complexity created by differences introduced by region, identity, or context cannot be fully achieved through statistics alone.

Policy Implications

This paper examines the relationship between language proficiency and wage premiums, with a particular focus on Lithuania, through a comparative lens that includes regions such as Québec, Catalonia, and Switzerland. By applying analytical models including TSLS, OLS, and PSM, and simulating data relevant to the Lithuanian labour market, the study highlights the significant economic value associated with multilingual competence.

The findings suggest that language proficiency is not merely a communicative asset but also a quantifiable economic resource, especially in multilingual labour markets. Regions that have institutionalised bilingualism or multilingualism tend to reward language skills more systematically, often through public sector wage premiums or formal recognition of language rights. In contrast, in Lithuania, where Lithuanian remains the sole official language and the labour market integration of minorities and migrants is governed by relatively rigid monolingual policies, wage premiums are likely to depend more heavily on Lithuanian language proficiency than on broader multilingual competence.

From a policy perspective, this analysis suggests that greater recognition of language diversity and more inclusive language policies could contribute to better labour market integration outcomes. Specifically, incorporating second language support programs, recognising the economic potential of Russian, Polish, or Ukrainian speakers, and aligning integration policy with labour market demand could reduce inequalities and increase economic efficiency. Moreover, improving the transparency of language requirements in job descriptions and formalising language premium structures could support fairer wage distributions.

While the use of simulated data limits direct empirical generalisation, the framework applied here provides a valuable foundation for future research using real Lithuanian labour force or PIAAC datasets. Further investigation could explore sectoral variation in wage premiums and longitudinal impact of language acquisition on income mobility.

In conclusion, language proficiency remains a critical yet under-recognized factor in wage dynamics in Lithuania. To build a more inclusive and

productive economy, policymakers must move beyond symbolic language policy and toward practical, data-informed strategies that acknowledge the real economic value of linguistic capital.

Ethical considerations

This study is a review with an illustrative simulation based on the public PIAAC variable structure. It uses no human subjects or identifiable data. All sources are publicly available and properly cited. Therefore, no ethics approval was required and no consent was sought.

Conclusions

This study aimed to systematically review and integrate previous research on the economic value of multilingualism, utilising various statistical models. The primary objectives were to: (1) review how statistical models have been applied to estimate wage premiums associated with multilingualism; (2) integrate these models into a unified analytical framework to improve coherence and comparability; and (3) identify gaps in existing research and suggest methodological improvements. The findings from the reviewed studies reveal important insights into how language proficiency influences economic outcomes across different multilingual labour markets.

Beyond individual wage effects, the proposed framework supports sustainable multilingualism by aligning language policy with labour demand, formalising transparent language criteria in hiring, and incentivising skills development across both majority and minority languages. Embedding these levers system-wide can sustain linguistic diversity and reduce access frictions, turning multilingualism into durable human-capital and productivity gains for Lithuania.

Addressing Objective 1: Application of Statistical Models

The review demonstrates that statistical models such as Two-Stage Least Squares (TSLS), Propensity Score Matching (PSM), Panel Data Analysis, and Fixed/Random Effects Models have been effectively applied to estimate

wage premiums associated with language proficiency. Each model addresses specific methodological challenges, enhancing the robustness of findings derived from cross-sectional and longitudinal data. For instance, TSLS has been particularly effective in addressing endogeneity by using instrumental variables, as demonstrated in Di Paolo & Raymond (2012) and Albouy (2008). PSM, applied by researchers such as Chiswick & Miller (1995) and Budría et al. (2019), effectively controls selection bias by creating matched samples of individuals with similar characteristics. Moreover, Panel Data Analysis allows researchers to capture dynamic changes over time, enhancing the ability to track long-term economic impacts of language skills (Aldashev et al., 2009; Ridala, 2020). Finally, Fixed and Random Effects Models provide valuable tools for controlling unobserved heterogeneity across regions, thereby enhancing cross-regional comparisons (Grin & Sfreddo, 1998; Cattaneo & Winkelmann, 2005).

Addressing Objective 2: Integrating Models into a Unified Framework

The integration of various statistical models into a comparative analytical framework provides a coherent structure for examining wage premiums associated with multilingualism. The framework demonstrates how TSLS, PSM, Panel Data Analysis, and Fixed/Random Effects Models are complementary rather than mutually exclusive. While TSLS addresses endogeneity concerns, PSM controls for selection bias, and Panel Data Analysis captures longitudinal effects. Moreover, Fixed and Random Effects Models allow researchers to make cross-regional comparisons, accounting for institutional and cultural factors.

The unified framework enhances the comparability of findings across diverse multilingual contexts by systematically integrating previous research methodologies. This approach allows for a more comprehensive understanding of how language skills influence economic outcomes, particularly when institutional factors such as Bill 101 in Québec and the official status of Catalan in Spain play a significant role. The findings confirm that institutional recognition of language skills positively influences wage premiums and employment opportunities for minority language speakers.

Addressing Objective 3: Identifying Gaps and Suggesting Improvements

While the application of statistical models has significantly improved the reliability of estimates related to the economic value of multilingualism, several limitations persist. First, the reliance on cross-sectional data in many studies limits the ability to capture long-term economic outcomes. Second, while TSLS effectively addresses endogeneity, its applicability is constrained by the availability of valid instrumental variables. Third, while PSM reduces selection bias, it cannot fully address unobserved heterogeneity, particularly when data availability is restricted.

Future research should focus on expanding longitudinal datasets, standardizing data collection methods, and integrating qualitative factors such as cultural integration, social capital, and perceived value of language skills. Additionally, the application of statistical models should be complemented by mixed-method approaches to provide a more nuanced understanding of how language skills influence socioeconomic outcomes. Addressing these limitations will enhance the robustness of findings and contribute to a more coherent understanding of multilingualism's economic value.

This study contributes to the literature by presenting a unified analytical framework that integrates statistical models applied to estimate wage premiums associated with language proficiency. By providing a coherent basis for cross-regional comparison, this framework enhances the robustness and comparability of findings across diverse multilingual contexts. The proposed framework offers valuable insights for policymakers seeking to improve economic integration and labour market accessibility in increasingly multilingual societies.

References

Albouy, D. (2008). *The wage gap between Francophones and Anglophones: A Canadian perspective, 1970–2000. Canadian Journal of Economics / Revue canadienne d'économique, 41*(4), 1211–1238. https://doi.org/10.1111/j.1540-5982.2008.00500.x

Aldashev, A., Gernandt, J., & Thomsen, S. L. (2009). The immigrant–native wage gap in Germany: Are East Europeans worse off? *Labour Economics, 16*(1), 36–48. https://doi.org/10.1016/j.labeco.2008.11.004

Becker, G. S. (1993). *Human capital: A theoretical and empirical analysis, with special reference to education* (3rd ed.). University of Chicago Press. https://doi.org/10.7208/chicago/9780226041223.001.0001

Budría, S., Colino, A., & Martínez de Ibarreta, C. (2019). The impact of host language proficiency on employment outcomes among immigrants in Spain. *Empirica, 46*(3), 625–652. https://doi.org/10.1007/s10663-018-9414-x

Cattaneo, M. A., & Winkelmann, R. (2005). Earning differentials between German and French speakers in Switzerland. *Swiss Journal of Economics and Statistics, 141*(II), 191–212. https://econpapers.repec.org/article/sesarsjes/2005-ii-2.htm

Chiswick, B. R., & Miller, P. W. (1995). *The endogeneity between language and earnings: International analyses. Journal of Labor Economics, 13*(2), 246–288. https://doi.org/10.1086/298374

Di Paolo, A., & Raymond, J. L. (2012). Language knowledge and earnings in Catalonia. *Regional Science and Urban Economics, 42*(5), 865–878. https://doi.org/10.1016/j.regsciurbeco.2012.04.003

Eurostat. (2019). *Foreign language learning statistics.* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Foreign_language_learning_statistics

Eurostat. (2023a). *Demographic statistics: Foreign-born population – Lithuania.* https://ec.europa.eu/eurostat

Eurostat. (2023b). *Demography of Europe – 2023 interactive edition.* https://ec.europa.eu/eurostat/web/interactive-publications/demography-2023

European Labour Authority (EURES). (2023). *Labour shortages and surpluses 2022: Report based on data from national public employment services.* https://eures.ec.europa.eu/publications/labour-shortages-and-surpluses-2022_en

Gazzola, M., & Mazzacani, D. (2019). Foreign language skills and employment

status of European natives: Evidence from Germany, Italy and Spain. *Empirica, 46*(4), 713–740. https://doi.org/10.1007/s10663-019-09460-7

Grin, F., & Sfreddo, C. (1998). Language-based earnings differentials on the Swiss labour market: Is Italian a liability? *International Journal of Manpower, 19*(7), 520–532. https://doi.org/10.1108/01437729810237196

Hahm, S., & Gazzola, M. (2022). The value of foreign language skills in the German labor market. *Labour Economics, 76*, 102150. https://doi.org/10.1016/j.labeco.2022.102150

Isphording, I. E. (2015). Language and labor market success. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., Vol. 13, pp. 260–265). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.94033-9

Lietuvos Respublikos socialinės apsaugos ir darbo ministerija. (2018). *Dėl užsieniečių teisės į socialinę integraciją ir darbo rinkos prieinamumą: Įsakymas Nr. A1-778.* https://www.e-tar.lt/portal/lt/legalAct/cc845690052311e9a5eaf2cd290f1944

Liwiński, J. (2019). The wage premium from foreign language skills. *Empirica, 46*(2), 251–284. https://doi.org/10.1007/s10663-019-09459-0

OECD. (2016). *The survey of adult skills: Reader's companion* (2nd ed.). OECD Publishing. https://doi.org/10.1787/9789264258075-en

Rendon, S. (2007). The Catalan premium: Language and employment in Catalonia. *Journal of Population Economics, 20*(3), 669–686. https://doi.org/10.1007/s00148-005-0048-5

Ridala, S. (2020). Language skills and labour market returns: A meta-regression analysis. *Language Problems and Language Planning, 44*(2), 200–241. https://doi.org/10.1075/lplp.20012.rid

Seimas of the Republic of Lithuania. (1995–2023). *Law on the State Language (Valstybinės kalbos įstatymas).* https://www.lrs.lt

Seimas of the Republic of Lithuania. (2004, April 29). *Law on the Legal Status of Aliens (No. IX-2206).* https://www.lithuanialaw.com/lithuanian-law-on-the-legal-status-of-aliens-526

Shapiro, D. M., & Stelcner, M. (1997). Language and earnings in Québec:

Trends over twenty years, 1970–1990. *Canadian Public Policy / Analyse de Politiques,* 23(2), 115–140. https://doi.org/10.2307/3551481

Stöhr, T. (2015). The returns to occupational foreign language use: Evidence from Germany. *Labour Economics, 32*, 86–98. https://doi.org/10.1016/j.labeco.2015.01.004

Toomet, O. (2011). Learn English, not the local language! Ethnic Russians in the Baltic States. *American Economic Review, 101*(3), 526–531. https://doi.org/10.1257/aer.101.3.526

World Economic Forum. (2023). *The future of jobs report 2023.* https://www.weforum.org/publications/the-future-of-jobs-report-2023

Zorlu, A., & Hartog, J. (2018). *The impact of language on socioeconomic integration of immigrants (IZA Discussion Paper No. 11485).* Institute of Labor Economics (IZA). https://www.iza.org/publications/dp/11485/the-impact-of-language-on-socioeconomic-integration-of-immigrants

Milda Kiškė
Vilniaus kolegija (VIKO), Lietuva
m.kiske@eif.viko.lt

KALBINĖS KOMPETENCIJOS IR DARBO UŽMOKESČIO SĄSAJOS: ANALIZĖS MODELIŲ APŽVALGA LIETUVOS DARBO RINKOS KONTEKSTE

Anotacija. Šiame straipsnyje, siekiant įvertinti daugiakalbystės ekonominę vertę Lietuvos darbo rinkai, apžvelgiami daugelis tarptautinių empirinių tyrimų ir siūloma lyginamoji analitinė sistema. Tyrimas remiasi žmogiškojo kapitalo teorijos bei kalbos ekonomikos ir komunikacijos efektyvumo modelių koncepcijomis. Apibendrinant pagrindinius daugiakalbės darbo rinkos tyrimų rezultatus, straipsnyje parodoma, kaip, taikant įvairius statistinius metodus, galima analizuoti darbo užmokesčio skirtumus, susijusius su kalbiniais gebėjimais. Tyrime naudojamas iliustratyvus imituotų duomenų rinkinys, pagrįstas PIAAC metodika. Duomenys atskleidžia, kad asmenys, mokantys valstybinę kalbą (priimančiosios šalies), gali tikėtis vidutiniškai 18,8 % didesnio darbo užmokesčio. Tyrimo išvados rodo: siekiant sustiprinti darbo jėgą Lietuvoje, būtų naudinga integruoti kalbos politiką į švietimo ir užimtumo sistemas. Šiame straipsnyje pateikti statistiniai rezultatai pagrįsti imituotais duomenimis, modeliuojamais pagal PIAAC sistemą, ir neatspindi empirinių išvadų, gautų iš realių duomenų. Be to, siūlomas analitinis modelis suteikia naują perspektyvą ir užpildo konkrečią spragą esamų tyrimų apie daugiakalbystės svarbą Lietuvos ekonomikai. Pabrėžiami tokie politikos svertai kaip skaidrūs kalbos reikalavimai, tikslinė parama antrosios kalbos įsisavinimui ir mažumų kalbų kapitalo pripažinimas, gali padėti įgyvendinti tvarią daugiakalbystę, t. y. išlaikyti kalbinę įvairovę, ir kartu sulyginti darbuotojų galimybes patekti į Lietuvos darbo rinką.

Pagrindinės sąvokos: darbo rinka; darbo užmokesčio priedas; daugiakalbystė; ekonominė vertė; kalbiniai gebėjimai.