Adult Education and Full-time Professionals' Problem Solving Skills: Insights From the Survey of Adult Skills

Author: Shiya Yi

Boston College Electronic Thesis or Dissertation, 2020

BOSTON COLLEGE
Lynch School of Education

Department of
Measurement, Evaluation, Statistics and Assessment

ADULT EDUCATION AND FULL-TIME PROFESSIONALS' PROBLEM SOLVING SKILLS: INSIGHTS FROM THE SURVEY OF ADULT SKILLS

Dissertation
by
SHIYA YI

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2020

ADULT EDUCATION AND FULL-TIME PROFESSIONALS' PROBLEM SOLVING SKILLS: INSIGHTS FROM THE SURVEY OF ADULT SKILLS

By
Shiya Yi
Dr. Henry I. Braun, Dissertation Chair

ABSTRACT

Sponsored by OECD, PIAAC represents the first attempt to assess adult problem solving in technology-rich environments (PS-TRE) on an international scale that is comparable cross-culturally and cross-nationally. The objectives of this study are to study (1) the distributions of PS-TRE proficiency scores across 14 selected countries and (2) within each country, the associations between PS-TRE proficiency scores and the different formats of adult education and training (AET) participation. Using data on full-time professionals (at least 25 years old) from these countries, propensity score weighting was applied to estimate the associations between the different formats of AET participation and their PS-TRE proficiency scores. To place these estimates in context, parallel analyses were conducted – one with the sample of full-time associates in the 14 selected countries and the other with full-time professionals' Literacy and Numeracy proficiency scores as measured by PIAAC.

The results showed that after controlling for socio-demographic background, occupational categories, use of key information-processing skills (both at home and at work), as well as use of generic workplace skills, no consistent pattern was found across the 14 selected countries. At the individual country level, scattered significant relationships were identified. For example, in Denmark, both formats of AET participation (vs. None) are significantly and positively associated with full-time professionals' PS-TRE proficiency scores and their probability of scoring in the top quartile of the PS-TRE distribution ($p < .01$). While in the United States, Formal AET (vs. None) is significantly and positively associated with full-time associates' PS-TRE proficiency scores and their probability of scoring in the top quartile of the PS-TRE distribution ($p < .01$).

The variations in relationships between the different formats of AET participation and working adults' skills proficiency across domains and samples indicate the necessity of conducting qualitative research on AET programs in individual countries. Furthermore, to provide recommendations tailored to the specific needs of each country, a fine-grained classification of AET programs based on the OECD guideline was suggested.

ACKNOWLEDGEMENTS

I thank first my advisor and dissertation chair Dr. Henry Braun, who has significantly shaped my statistical thinking and this work. I learn from you always. I thank my readers Dr. O'Dwyer, Dr. Martin and Dr. Kirsch who have also generously contributed to this dissertation.

I thank my dearest, most amazing and God-loving Mama. You are my best friend. Your unconditional love, trust and support suffuses my every step, big and small. Home is when I am with you. I dedicate this dissertation to you.

I thank Dr. Hesse-Biber for her unfailing support. I thank Dr. Piselli for listening and counseling. I thank Mr. and Mrs. Foote, your kindness and wisdom are immensely appreciated. I hope I'll be able to return the favor to you sometime.

I thank my lovely friends at Harvard Medical School for being there when I needed inspiration. It's a pleasure to know every single one of you and I thank God for our time together. I thank the faculty of MIT Critical Data for providing training in data science and medicine, the two subjects that pull at my heart strings.

Last but not least, I thank myself. This dissertation was written amid my personal turmoil. I did not give up and it is now done.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS V

LIST OF TABLES IX

LIST OF FIGURES XIV

LIST OF EQUATIONS.. XVII

CHAPTER 1. INTRODUCTION 1

1.1 BACKGROUND 1

1.2 PIAAC OVERVIEW 6

1.3 SUBJECTS OF INTEREST 14

1.4 RESEARCH OBJECTIVES AND RESEARCH QUESTIONS 15

1.5 SIGNIFICANCE OF THE STUDY 18

1.6 LIMITATIONS OF THE STUDY 20

1.7 DISSERTATION ORGANIZATION (ROADMAP) 21

CHAPTER 2. LITERATURE REVIEW 23

2.1 THEORETICAL FRAMEWORK FOR FACTORS RELATED TO PROBLEM SOLVING IN TECHNOLOGY-RICH ENVIRONMENTS (PS-TRE) 23

2.2 CHALLENGES IN DRAWING CAUSAL INFERENCES FROM OBSERVATIONAL STUDIES 38

2.3 PROPENSITY SCORE METHODS TO REDUCE BIAS 42

2.4 STRENGTHS AND LIMITATIONS OF PROPENSITY SCORE METHODS 54

CHAPTER 3. RESEARCH DESIGN & METHODS... 55

3.1 SAMPLE OF FULL-TIME PROFESSIONALS FROM THE 14 SELECTED COUNTRIES............. 55

3.2 SELECTION OF COVARIATES 58
3.3 RESEARCH QUESTIONS 64
3.4 DESCRIPTION OF FULL-TIME PROFESSIONALS SAMPLE IN THE 14 SELECTED COUNTRIES 92
3.5 DESCRIPTION OF FULL-TIME ASSOCIATES SAMPLE IN THE 14 SELECTED COUNTRIES 114

CHAPTER 4. RESULTS 138

4.1 RESEARCH QUESTION ONE 138
4.2 RESEARCH QUESTION TWO 158
4.3 RESEARCH QUESTION THREE 175
4.4 RESEARCH QUESTION FOUR 208

CHAPTER 5. CONCLUSIONS 235

5.1 OVERVIEW OF RESEARCH QUESTIONS 235
5.2 SUMMARY OF FINDINGS BY RESEARCH QUESTION 237
5.3 LIMITATIONS AND FUTURE WORK 245

REFERENCES 249

APPENDIX A. PARTICIPATING COUNTRIES IN PIAAC (ROUND 1) AND THEIR SAMPLE SIZES 259

APPENDIX B. PERCENTAGES OF RESPONDENTS TAKING DIFFERENT PATHWAYS IN PIAAC 260

APPENDIX C. INDICATORS OF SKILLS USE AT WORK AND SAMPLE TASKS 261

APPENDIX D. ACCOUNTING FOR IMPUTATION ERROR VARIANCE COMPONENT 262

APPENDIX E. CORRELATION MATRIX BETWEEN SKILLS USED AT WORK FOR FULL-TIME PROFESSIONALS.. 263

APPENDIX F. CORRELATION MATRIX BETWEEN SKILLS USED AT WORK FOR FULL-TIME ASSOCIATES.. 264

APPENDIX G. RESULTS FROM ONE-WAY ANOVA POST-HOC TESTS 265

LIST OF TABLES

Table 1. A Two-by-Two Matrix of Workplace Tasks 1

Table 2. Overview of Indicators of Skills Use at Work... 34

Table 3. Potential Outcomes in the Presence and Absence of a Treatment.................................. 43

Table 4. The ISCO-08 Framework & Its Occupational Classifications of "Professionals" 56

Table 5. Summary of Sample Size Procedures by Country (absolute number & percentage) 58

Table 6. Overview of Predictor Variables Entered into the Logistic Regression Models 61

Table 7. Binary Coding for Socio-Demographic Background Factors (Model 1) 75

Table 8. Example from the Japanese Results of Multinomial Logistic Regression 78

Table 9. Observed Pre-Treatment Covariates for Propensity Score Estimation........................... 87

Table 10. Estimates of Average Treatment Effects (ATEs) for Three Treatment Levels 90

Table 11. Classification of AET Participation Rates, by Country 112

Table 12. The Rankings of Countries by the Mean PS-TRE Proficiency Score for Full-time Professionals 139

Table 13. The Rankings of Countries by the Mean PS-TRE Proficiency Score for Full-time Associates (Table 12b, corresponding to Table 12)... 153

Table 14. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Belgium 160

Table 15. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – the Slovak Republic 161

Table 16. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – the Czech Republic.. 161

Table 17. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Ireland 161

Table 18. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Japan 162

Table 19. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – South Korea 163

Table 20. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Denmark 163

Table 21. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Germany 164

Table 22. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – The Netherlands 164

Table 23. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – The United Kingdom 164

Table 24. Distribution of Significant Indicators of Skills Use across Countries – Non-formal vs. None 165

Table 25. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Czech Republic 167

Table 26. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Denmark 168

Table 27. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Belgium 168

Table 28. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Sweden 168

Table 29. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the United States 168

Table 30. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Japan 169

Table 31. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – South Korea 169

Table 32. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Slovak Republic 170

Table 33. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Czech Republic 172

Table 34. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Netherlands 172

Table 35. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Norway 172

Table 36. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the United Kingdom 172

Table 37. Distribution of Significant Indicators of Skills Use across Countries – Formal vs. None AET Participation 174

Table 38. Summary of the Estimated GPS Matrix for Full-time Professionals by Their Actual AET Participation by Country 177

Table 39. Goodness of Fit of Multinomial Logistic Regression Models by Country 181

Table 40. Predictive Power of Multinomial Logistic Regression Models by Country.............. 183

Table 41. Summary of the Normalized Weights by Full-time Professionals' Actual AET Participation by Country 192

Table 42. Summary of the Standardized Effect Size between Pairs of AET Participation Formats among Full-time Professionals by Country 196

Table 43. Countries with Maximum Standardized Effect Sizes > .4 between None and Formal AET Participation 199

Table 44. Countries with Maximum Standardized Effect Sizes > .4 between Non-formal and Formal AET Participation... 200

Table 45. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' PS-TRE Scores... 204

Table 46. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the PS-TRE Score Distribution... 207

Table 47. Summary of the Estimated GPS Matrix for Full-time Associates by Their Actual AET Participation by Country (Table 38b, corresponding to Table 38) 209

Table 48. Summary of the Normalized Weights by Full-time Associates' Actual AET Participation by Country (Table 41b, corresponding to Table 41) 217

Table 49. Summary of the Standardized Effect Size between Pairs of AET Participation 221

Table 50. Countries with Maximum Standardized Effect Sizes > .4 between None and Formal AET Participation (Table 43b, corresponding to Table 43) 223

Table 51. Countries with Maximum Standardized Effect Sizes > .4 between Non-formal and Formal AET Participation (Table 44b, corresponding to Table 44).. 224

Table 52. Estimated ATEs of the Different Formats of AET Participation on Full-time Associates' PS-TRE Scores (Table 45b, corresponding to Table 45) 225

Table 53. Estimated ATEs of the Different Formats of AET Participation on Full-time Associates' Probability of Scoring in the Top Quartile of the PS-TRE Score Distribution (Table 46b, corresponding to Table 46) 229

Table 54. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the LIT Distribution 231

Table 55. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' NUM Proficiency Scores 233

Table 56. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the NUM Distribution 233

Table 58. Summary of Estimated ATEs of the Different Formats of AET Participation on the Different Outcomes and Different Samples by Country 243

LIST OF FIGURES

Figure 1. The Distribution of Full-time Professionals (≥ 25 years old) by Country 93

Figure 2. Age Distribution of Full-time Professionals (≥ 25 years old) by Country 95

Figure 3. Gender Distribution of Full-time Professional (≥ 25 years old) by Country 97

Figure 4. Education Distribution of Full-time Professionals (≥ 25 years old) by Country 98

Figure 5. Parental Ed. Distribution of Full-time Professionals (≥ 25 years old) by Country 101

Figure 6. Occupation Distribution of Full-time Professionals (≥ 25 years old) by Country 102

Figure 7. Distribution of the Use of Key Information-Processing Skills at Work 105

Figure 8. Distribution of the Use of Key Information-Processing Skills at Home..................... 107

Figure 9. Distribution of the Use of Workplace Skills/Job-related Activities 109

Figure 10. Distribution of AET Participation among Full-time Professionals (≥ 25 years old) by Country 114

Figure 11. The Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 1b, corresponding to Figure 1) 115

Figure 12. Age Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 2b, corresponding to Figure 2) 117

Figure 13. Gender Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 3b, corresponding to Figure 3) 119

Figure 14. Education Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 4b, corresponding to Figure 4) 120

Figure 15. Parental Ed. Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 5b, corresponding to Figure 5) 123

Figure 16. Occupation Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 6b, corresponding to Figure 6) 124

Figure 17. Distribution of the Use of Key Information-Processing Skills at Work (Figure 7b, corresponding to Figure 7) 126

Figure 18. Distribution of the Use of Key Information-Processing Skills at Home (Figure 8b, corresponding to Figure 8) 129

Figure 19. Distribution of the Use of Workplace Skills/Job-related Activities (Figure 9b, corresponding to Figure 9) 132

Figure 20. Distribution of AET Participation among Full-time Associates (≥ 25 years old) by Country (Figure 10b, corresponding to Figure 10) 137

Figure 21. The Distributions of Full-time Professionals PS-TRE Proficiency Scores by Country 139

Figure 22. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Gender and Country 142

Figure 23. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Age and Country 143

Figure 24. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Gender, Age and Country 145

Figure 25. Scatterplot of the Mean PS-TRE Scores vs. the Mean Literacy Scores across the 14 Countries 147

Figure 26. Scatterplot of the Mean PS-TRE Scores vs. the Mean Numeracy Scores across the 14 Countries 148

Figure 27. Fisher's Transformation of the Correlation between PS-TRE and Literacy 149

Figure 28. Fisher's Transformation of the Correlation between PS-TRE and Numeracy.......... 149

Figure 29. The Distribution of Full-time Associates PS-TRE Proficiency Scores by Country (Figure 21b, corresponding to Figure 21).. 151

Figure 30. Distribution of Full-time Associates PS-TRE Proficiency Scores by Gender, Age and Country (Figure 24b, corresponding to Figure 24)... 155

Figure 31. Fisher's Transformation of the Correlation between PS-TRE and Literacy (Figure 27b, corresponding to Figure 27).. 157

Figure 32. Fisher's Transformation of the Correlation between PS-TRE and Numeracy (Figure 28b, corresponding to Figure 28).. 157

Figure 33. Overview of Full-time Professionals' Participation in the Different Formats of AET Programs by Country 159

Figure 34. Distribution of Different Formats of AET Participation by Age Cohort in Czech ... 167

Figure 35. Distribution of Different Formats of AET Participation by Occupational Category in the Slovak Republic 171

Figure 36. Kernel Density Plots of the Distribution of the Estimated GPS for Non-formal, Formal and None AET Participation.. 187

Figure 37. Overview of Full-time Associates' Participation in the Different Formats of AET Programs by Country (Figure 33b, corresponding to Figure 33)... 208

Figure 38. Kernel Density Plots of the Distribution of the Estimated GPS for Non-formal, Formal and None AET Participation (Figure 36b, corresponding to Figure 36) 212

LIST OF EQUATIONS

Equation 1 42

Equation 2 43

Equation 3 43

Equation 4 43

Equation 5 46

Equation 6 82

Equation 7 82

Equation 8 83

Equation 9 83

Equation 10 83

Equation 11 83

Equation 12 88

Equation 13 89

Equation 14 210

Equation 15 210

Equation 16 211

Equation 17 211

Chapter 1. Introduction

1.1 Background

Over the past 30 years, the rapid advance of computers and related technology has not only revolutionized our access to information and communication capabilities, but also profoundly transformed job tasks and labor market structures worldwide. In developed countries particularly, computers are taking over routine cognitive and manual tasks that can be accomplished by following explicit rules and leaving workers to accomplish non-routine cognitive tasks "demanding flexibility, creativity, generalized problem-solving, and complex communications" (Bresnahan, 1999). In an empirical exploration of the skill content of recent technological change, the authors referred to "nonroutine" tasks as those "for which the rules are not sufficiently well understood to be specified in computer code and executed by machines (Autor, Levy & Murnane, 2003). Table 1 provides examples of workplace tasks (routine vs. nonroutine, manual vs. information processing) and hypothesizes the impact of computerization for each category.

Table 1. A Two-by-Two Matrix of Workplace Tasks

	Routine Tasks	**Nonroutine Tasks**
Manual Tasks	• repetitive assembly • picking or sorting **(Substantial Substitution)**	• truck driving • janitorial services **(Strong Complementarities)**
Analytic & Interactive Tasks	• repetitive customer service • calculation • record-keeping **(Substantial Substitution)**	• medical diagnosis • legal writing • managing others • persuading/selling • forming/testing hypotheses **(Limited Opportunities for Substitution or Complementarity**

In response to the evolution of job skill demands favoring more technology-savvy workers in today's increasingly competitive labor markets, many education and training systems have extended their focus from traditional cognitive skills (i.e., Literacy and Numeracy) to comprehensive problem solving skills and more broadly, the so-called 21st century skills.

21st Century Skills and Complex Problem Solving

Despite the lack of definitive research on the range of skills that have been associated with the overarching label "21st century skills", the National Research Council committee (2012) took initial steps toward classifying the various terms for "21st century skills" into three broad domains of competence – cognitive, intrapersonal and interpersonal:

- The Cognitive Domain includes competencies such as information literacy, nonroutine problem solving, reasoning and argumentation, critical thinking and innovation.
- The Intrapersonal Domain includes competencies such as flexibility, initiative, appreciation for diversity and metacognition (the ability to reflect on one's own learning and make adjustments accordingly)
- The Interpersonal Domain includes competencies such as communication, collaboration, responsibility and conflict resolution.

Although having identified complex/nonroutine problem solving as cognitive competencies along with verbal and quantitative literacy, the NRC committee (2012) recognized that "there are areas of overlap between and among the individual '21st century skills' and the larger competency clusters" and that complex/nonroutine problem solving involves the interplay of competencies across all three domains. The ability to solve problems is first considered as a sequence of complex cognitive operations (Newell & Simon, 1972). In order to solve a problem, individuals have to understand the nature of the problem ("problem finding"), define a set of sub-goals and steps through which the problem may be solved ("problem shaping") and perform the actions required to reach those sub-goals until the situation reaches a satisfactory state ("problem solving") (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c). Throughout the process, individuals have to apply intrapersonal skills (also called "metacognition") to reflect on whether a problem solving strategy is working and, where necessary, to switch to another strategy accordingly (Levy & Murnane, 2004). Last but not least, problem solving in concrete everyday situations often involves other persons and thus interpersonal skills (e.g., communication and collaboration) are essential for assessing problem solving skills (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c).

Importance of Problem Solving Skills in the Workplace

Residing at the intersection of cognitive, intrapersonal and interpersonal competencies, problem solving has received heightened attention from both the world of education and the world of work. As early as 2003, the Programme for International Student Assessment (PISA)[1] has included an assessment of students' problem solving skills since 2003 (OECD, 2003) and introduced collaborative problem solving in 2015 (OECD, 2015, 2017). More recently, the National Center for Education Statistics (NCES) has been actively considering adding an

[1] PISA is a triennial international survey sponsored by the Organisation for Economic Co-operation and Development (OECD). First conducted in 2000, it measures 15-year-old students' scholastic performance on science, mathematics, reading, collaborative problem solving and financial literacy (OECD, 2003, 2015, 2017).

assessment of collaborative problem solving to the National Assessment of Educational Progress (NAEP)[2] (NCES, 2017). These efforts resonate closely with voices from labor markets.

In the current fast changing global economy, employers identify problem solving as crucial to the success of their organizations. The Partnership for 21st Century Skills (2010) argued that student success in college and careers requires problem solving skills. The NRC report (2012) held that problem solving skills "are important to success in education, work, and other areas of adult responsibility". In a 2010 survey conducted by the American Management Association, 70% of 2000+ business leaders identified problem solving skills as one of priorities when hiring and evaluating employees (AMA, 2010). Similarly, Casner-Lotto and Barrington (2006) found that among 400 surveyed employers, 92.1% identified problem solving as a very important skill for 4-year college graduates to be successful in today's workforce.

The majority of business surveys in the literature rated problem solving skills so highly because the core ingredients are relevant to virtually any job role in the 21st century. Specifically, effective problem solving skills enable employees to (1) use available resources to resolve issues in a constructive manner and (2) work more efficiently with co-workers, customers, partners and suppliers (Casner-Lotto & Barrington, 2006; AMA, 2010; Partnership for 21st Century Skills, 2010; NRC, 2012). At the individual level, as the job market becomes increasingly competitive, problems solving is the key to compete for improved employment opportunities and higher wages (Hodge & Lear, 2011).

[2] NAEP, widely known as the Nation's Report Card, is the largest nationally representative and continuing assessment of what U.S. students know and can do in a variety of subject areas. First administered in 1969, it periodically measures fourth-, eighth- and twelfth-graders' performance on science, mathematics, reading, writing, arts, civics, economics, geography, U.S. history, and in Technology and Engineering Literacy (TEL) (NCES, 2010, 2013, 2017).

PISA's Definition of Problem Solving Competencies

As the first direct assessment of problem solving competencies that apply across different school subject areas, the PISA 2003 defined problem solving as an individual's capacity to "understand problems situated in novel and cross-curricular settings, identify relevant information or constraints, represent possible alternatives or solution paths, develop solution strategies, solve problems, check or reflect on the solutions and communicate the solutions" (OECD, 2003). However, the comprehensive (i.e., triple-strand) nature of complex problem solving presents a challenge for consistent and accurate measurement across traditional and technological contexts (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c).

From a practice perspective, complex problem solving usually requires some sort of tools. Paper and pencil are examples of tools traditionally used to process information. However, in today's technology-rich societies, it has become common for people to rely on computers and computer networks (e.g. web-based services and desktop software) to solve their problems. Although emerging technologies are meant to facilitate problem solving, they may also make the problem more difficult, "especially when a person has limited knowledge and experience with the use of those tools and technologies" (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c). Therefore, the integration of computer literacy and problem solving skills has been considered as potentially high return education and training investments (Lazonder & Rouet, 2008). In effort to contribute to these important undertakings, the Programme for the International Assessment of Adult Competencies (PIAAC) developed and conducted the first-ever direct assessment of problem solving in technology-rich environments (PS-TRE), targeting adults from ages 16 to 65 (OECD, 2013c).

1.2 PIAAC Overview

Data from the Survey of Adult Skills (PIAAC)

This study will use data from the first round of the Program for the International Assessment of Adult Competencies (PIAAC). Intended to be the most comprehensive initiative to collect large-scale observational data of adult skills that are comparable cross-culturally and cross-nationally, PIAAC (round 1) was administered to about 166,000 individuals aged 16 through 65 in 24 countries -- 22 OECD member countries plus Russian Federation and Cyprus -- mostly during the period of August 2011 to March 2012 (OECD, 2013c, p. 25). For each participating country, a minimum sample size of 5,000 completed cases was required if all three skill domains were assessed[3] in only one language (OECD, 2013b, p. 54). A number of participating countries chose to over-sample particular subgroups of the target population so that they could "obtain more precise estimates of proficiency by geographic area or for certain population groups" (OECD, 2013b). For example, Canada over-sampled persons aged 16-25, linguistic minorities, aboriginal population, and recent immigrants, as well as subgroups at the provincial and territorial level. Please refer to Appendix A for a complete list of sample sizes observed in the 24 countries, groups oversampled, cognitive domains assessed as well as assessment language(s).

The standard administration plan for PIAAC was for respondents to complete a background questionnaire before undertaking the cognitive assessment on a laptop computer. However, for respondents who (1) had no (or extremely limited) computer experience, (2) failed the

[3] Participating countries had the choice of assessing all three domains or assessing Literacy and Numeracy only. Assuming the assessment was administered in only one language, a minimum of 4,500 completed cases was required if only Literacy and Numeracy were assessed (OECD, 2013b, p. 54).

information and communications technology (ICT) core assessment or (3) opted out of taking the computer-based assessment, a paper-and-pencil version was available for Literacy and Numeracy only (OECD, 2013c). In total, 79.1% of the PIAAC sample attempted the ICT core and 73.6%[4] passed and continued on to take two assessment modules in either one or two of the three assessment domains – a combination of a Literacy and a Numeracy module, a combination of a Problem Solving and a Literacy or Numeracy module, or two Problem Solving modules. Some 24.4%[5] of respondents attempted the paper-based assessment (PBA) core and 21.4%[6] passed and went on to the full paper-based assessment – i.e., a literacy or numeracy module plus reading components. There was a small proportion of respondents (1.4%) for whom no background information was available. (OECD, 2013b, p. 50). The proportions of respondents in the full sample taking different pathways through the assessment are presented graphically in Appendix B. Since PS-TRE proficiency is the primary outcome of interest for this study, respondents following the paper-based pathway will not be included in the analysis.

All participating countries were required to administer the Literacy and Numeracy components of the assessments; however, administration of the PS-TRE and reading components assessments was optional. As it turned out, all but four countries (Cyprus, France, Italy and Spain) administered the computer-based PS-TRE assessment and all but three (Finland, France and Japan) administered the reading components assessment (OECD, 2013b, p. 51). For this

[4] 4.9% of the sample failed the ICT core Stage 1 and were directed to the paper-based assessment core; 0.6% filed the ICT core Stage 2 and were assigned reading components only (OECD, 2013b, p. 50).
[5] 9.3% of the sample had no prior computer experience, 10.2% opted out of the computer-based assessment and 4.9% attempted the ICT core Stage 1 but failed. These respondents (24.4%) were directed to the paper-based assessment core (OECD, 2013b, p. 50).
[6] A small proportion of respondents (1.8%) failed the PBA core and were assigned reading components only; another small proportion (1.2%) had no assessment data because they were unable or unwilling to undertake the assessment in the test language or languages available (OECD, 2013b, p. 50).

specific study, only countries that participated in the problem-solving assessment will be included (OECD, 2013b, p. 51).

A major component of the Survey of Adult Skills (PIAAC) is the task-based assessments of information-processing skills in three cognitive domains: literacy (Lit), numeracy (Num), and problem solving in technology-rich environments (PS-TRE). In addition to taking initiative in defining the PS-TRE conceptual framework (detailed Section 1.2), PIAAC built upon and expanded earlier conceptions of traditional cognitive skills from IALS and ALL to include a broader range of skills relevant to the information age. Specifically, PIAAC defined literacy and numeracy as (OECD, 2013c):

• Literacy: The ability to understand, evaluate, use and engage with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential. In particular, the skills of reading in digital environments were added to the PIAAC framework (PIAAC Literacy Expert Group, 2009a).

• Numeracy: The ability to access, use, interpret, and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life. This definition of numeracy is complemented with "numerate behavior" pertaining to four facets: contexts, responses, mathematical content/information/ideas and representations. (PIAAC Numeracy Expert Group, 2009b).

Demanding from both a cognitive and technological point of view, the PIAAC assessment battery enables researchers and policy makers to distinguish respondents with high cognitive skills according to whether they are able to apply their literacy/numeracy skills in tasks requiring high technological competence (OECD, 2010).

Results from the assessment were reported along three proficiency scales (representing Literacy, Numeracy and PS-TRE respectively), each ranging from 0 to 500. To facilitate interpretations of the scores, each scale was split into several ordered proficiency levels based on difficulty of "the knowledge and skills required to complete the tasks within those levels" (OECD, 2013a, Ch. 21, p. 3). Please refer to Chapter 21 of the *Technical Report of the Survey of Adult Skills (PIAAC)* (OECD, 2013a) for a detailed description of the content definition for each proficiency level per cognitive domain. The Literacy and Numeracy proficiency scales are defined in terms of six levels; while the PS-TRE scale is split into four levels. Respondents scoring at a particular level are also proficient at lower levels.

Besides assessing adults' proficiencies in key information-processing skills, the Survey of Adult Skills (PIAAC) collected a comprehensive set of data on basic socio-demographic background factors such as age, gender, family background and educational attainment, employment status and occupational category. What is more, a considerable proportion of the PIAAC background questionnaire was devoted to measuring the types and intensities of skills use at work through "the frequency with which individuals carry out a number of skills-related tasks in the context of their job" (OECD, 2016, p. 960). Specifically, this task-based or job-requirement approach (JRA) asked respondents "about the importance of different types of tasks performed at work and subsequently inferring the types of skills that are required from their answers" (OECD, 2013c). By focusing on job tasks, the JRA approach enables us to distinguish skills on the demand side of labor markets from skills on the supply side (i.e., the stock of skills of the population). Although self-report bias (e.g., recall bias) may still undermine the accuracy of respondents' answers, measures of skills use at work through the JRA module is considered to be more objective "in reflecting job-specific demands and skills use in the workplace" (OECD, 2016, p. 104) than older approaches relying on respondents' subjective self-assessments of the type and level of skills they possess[7].

In fact, the Field Test of PIAAC assessed the reliability of out-of-work respondents' ability to recall the activities of their most recent job in the previous 12 months. No indications were found that there was a serious recall bias. As a complement to the direct assessments in PIAAC, the JRA module not only opens a window into other important workplace skills[8], but also provides information needed to inform the heated debate on the mismatch between workers' "own skills" (skills brought to the workplace by employees) and "job skills" (skill demanded by employers) (OECD, 2013a, Ch. 3, p. 19-20).

By virtue of its sophisticated design, PIAAC can serve as a rich data source for comprehensive comparative analyses of skills maintenance and development across countries. The JRA module has made it uniquely suited to investigating the relationships between full-time professionals' proficiencies in key information-processing skills and a range of work-related experiences -- from the types and intensities of skills use at work to participation in formal and non-formal adult education and training (AET) programs -- while accounting for socio-demographic background factors and occupational categories. With due caution against making causal inferences from observational data, the findings can inform research and policy making

[7] There are at least two problems undermining the reliability of respondents' self-assessments of their skills. First, not all of them are sufficiently well-informed to report the type of skills involved in doing the job; this is especially true for young professionals who just held the job for a short time. Second, there seems to be a tendency for individuals to "talk up" their skill levels in order to boost their self-esteem; they are less likely to do so when reporting their activities rather than reporting how good they are in performing these activities (OECD, 2013a, Ch. 3, p. 20).

[8] There is a widespread feeling, supported by some case studies, that other skills are becoming increasingly relevant in modern workplaces. There is also evidence that some of these skills, like computing skills, are being rewarded in the labor market over and above the returns to the education that people had received (Dickerson & Green, 2004).

with respect to (1) the (apparent) effectiveness of education and training systems in preparing members of society for productive labor force participation and (2) the (apparent) success of workplace practice in maintaining and developing working adults' skills proficiency over a lifetime.

Last but not least, caution is called for because PIAAC, like many other large-scale assessment surveys, collected data at a single point in time (i.e. cross-sectional). Although statistical analyses based on large observational data from PIAAC can inform the relationships between adults' proficiency in processing information in technology-rich environments and their work-related experiences, researchers and policy makers need to confront a number of challenges as elaborated in Section 2.2.2.

PIAAC's Definition of Problem Solving in Technology-Rich Environments (PS-TRE)

Under the auspices of the Organization for Economic Cooperation and Development (OECD), the Programme for the International Assessment of Adult Competencies (PIAAC) is an innovative addition to previous large-scale assessment surveys of adults such as IALS and ALL[9]. As mentioned above, PIAAC represents the first attempt to assess problem solving on an international scale and as a single dimension. Unlike other traditional problem solving assessments, the PS-TRE domain of PIAAC is intended to "cover the specific class of problems people deal with when using information and communication technologies (ICT)" (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009). More specifically, PIAAC defined PS-TRE as follows (OECD, 2013c, p. 59):

---

[9] At international level, there were two large-scale adult surveys preceding PIAAC which was first administered in 2012: International Adult Literacy Survey (IALS, administered in 21 countries between 1994 and 1998) and Adult Literacy and Lifeskills (ALL, administered in 13 countries between 2003 and 2007). They were among the first comparative assessments designed to measure a range of adult skills in developed countries since the early 1990s.

"Problem solving in technology-rich environments involves using digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks. The first PIAAC problem solving survey will focus on the abilities to solve problems for personal, work and civic purposes by setting up appropriate goals and plans, accessing and making use of information through computers and computer networks."

This operational definition of PS-TRE reflects PIAAC's emphasis on the use of technology to solve problems and accomplish complex tasks through "information access, evaluation, retrieval and processing" (OECD, 2010). Comprising tasks of varying levels of difficulty from both a technological and cognitive point of view, the PS-TRE domain of PIAAC may be organized along three core dimensions (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c):

- Cognitive dimensions involve mental structures and processes by which a person actually performs problem solving. For example, goal setting and progress monitoring, planning and self-organizing, acquiring and evaluating information, and making use of information.
- Technologies include hardware devices (laptop computers for PIAAC), software applications and other functionalities through which problem solving is conducted.
- Tasks refer to circumstances that trigger a person's awareness and understanding of the problem, and that determine the actions to be taken in order to solve the problem. For example, questions and instructions presented to test takers, as well as the specific materials and time constraints associated with the test.

Although relying on the same “core” cognitive processes as the constructs of literacy and numeracy[10], the PS-TRE domain of PIAAC is distinct in that (1) it “specifically assesses goal setting, monitoring, and planning in technology-rich environments” and (2) “tasks will be carried out in environments that involve multiple, complex sources of information”. Numeracy and literacy demands were kept to a minimum “to increase the specificity and validity” of PS-TRE tasks. What is more, the PIAAC PS-TRE domain is different from the general ICT[11] domain because of its focus on the cognitive dimensions of problem solving, rather than “purely instrumental skills related to the knowledge and use of digital technologies” (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009c). Lastly, The PIAAC Expert Group (2009c) acknowledged there is more to “technology-rich environments” than merely “computers and computer networks” and a full assessment of PS-TRE would require a range of devices that mimic the diversity and versatility of digital technologies in today’s world. While setting the stage for further rounds of surveys, the first PIAAC assessment of PS-TRE was “limited to problems requiring the use of computers and Internet-based services” for feasibility reasons.

On a continuum ranging from 0 to 500, PS-TRE scores were split into three proficiency levels (with an additional category, “below Level 1”), while Literacy and Numeracy scores were split into five proficiency levels (with an additional category, “below Level 1”). Since PIAAC measures PS-TRE differently than Literacy and Numeracy, a simplified three-level proficiency scale was derived by Holzer and Lerman (2015) to facilitate comparisons between PS-TRE and

[10] For instance, the ability to decode printed symbols and at least a minimal working memory capacity are required for tasks in any of these domains (PIAAC Expert Group in Problem Solving in Technology-Rich Environments, 2009).
[11] ICT skills may be broadly defined as “the interest, attitude, and ability of individuals to appropriately use digital technology and communications tools” (Lennon et al., 2003).

traditional cognitive skills: Low, Medium and High Proficiency. In the PIAAC technical report (OECD, 2013a), PS-TRE score-point ranges are linked to discrete proficiency levels: below Level 1/"at-risk" (0-240); Level 1/"weak" (241-290); Level 2/"moderate" (291-340); Level 3/"strong" (341-500). The simplified three-level scale followed the same cut-off rules for PS-TRE proficiency levels but combined Level 1 and below into "low proficiency" for PS-TRE. Literacy and Numeracy share the same score-point range for proficiency levels: below Level 1 (0-175); Level 1 (176-225); Level 2 (226-275); Level 3 (276-325); Level 4 (326-375); Level 5 (376-500). The simplified three-level scale re-classified the measured skills levels into "low proficiency" (levels 2 and below), "medium proficiency" (level 3), or "high proficiency" (levels 4-5). These "slightly more stringent measures of proficiency (compared to the OECD)" have been shown to produce consist results with other analyses of the PIAAC data (Holzer & Lerman, 2015) indicating U.S. proficiency is low relative to other countries (OECD, 2013c, 2013e). In this study, these measures of proficiency will be applied to each of the 14 selected countries and the distribution of percentages across the three levels will be obtained for each key information-processing skill.

1.3 Subjects of Interest

The subjects of interest of this study are "full-time professionals" -- respondents who reported their current status as "full-time employed (self-employed or employee)" in one of the six occupational sub-categories that fall under the heading of "Professionals" as defined by the International Standard Classification of Occupations (ISCO, 2008). Note that the ISCO-08 occupational codes correspond closely to the standard 1- and 2- digit coding scheme used by the Bureau of Labor Statistics in the U.S.

Full-time professionals were chosen as the subjects of interest for four main reasons. First and foremost, they are the core component of the 21st century labor force. The levels and distributions of their skill sets directly shape the current and future landscape of a national economy. Second, they are the most privileged group with regard to human capital (e.g., educational attainment and socio-economic status). A scan of the PIAAC sample reveals that the majority of full-time professionals have obtained at least a bachelor's degree and all of them are currently working at high-skilled occupations. At the forefront of labor markets worldwide, they are expected to be equipped with the most advanced skills and competencies for the "21st century jobs" (i.e. jobs that require the 21 century skills). Third, they tend to be more aware of job-related education and training and are more likely to improve their already-high skills through such opportunities (OECD, 2013c). Moreover, better access to educational and socio-economic resources at the personal and/or organizational level often helps them find the appropriate program that meets their own training needs. Fourth, since most full-time professionals will stay in the labor force for the next few decades, understanding their trajectories of skill maintenance and development will (1) guide the integration of best work practices into current job tasks composition and (2) improve the design of adult education and training (AET) programs to better cater for the unique needs of this specific group.

1.4 Research Objectives and Research Questions

Notwithstanding the increased emphasis on problem solving in technology-rich environments (PS-TRE) both as a desirable educational outcome and as an important employability skill in today's digital world, little has been known with respect to full-time professionals regarding (1) the distributions of PS-TRE proficiency scores in different countries and (2) the associations between PS-TRE proficiency scores and different formats of participation in adult education and training (AET) programs.

The primary purposes of this study are to (1) describe full-time professionals' levels of PS-TRE proficiency scores and compare distributions of proficiency levels both within and across countries and (2) investigate the role of work-related experiences in accounting for variations in PS-TRE proficiency scores. Of particular interest will be full-time professionals' participation in three different formats of AET programs: None, Non-formal and Formal[12]. Investigating the relationships of participation in these AET programs and full-time professionals' problem solving skills in technology-rich environments at the individual country level will inform policy and practices regarding adult learning.

Accordingly, four research questions will be addressed in this study:

RQ1a. What are the distributions of full-time professionals' PS-TRE proficiency scores across the 14 selected countries that participated in the first round of PIAAC?

RQ1b. Within each country, how do the distributions of full-time professionals' PS-TRE proficiency scores vary by gender and age?

[12] Two items regarding AET participation were derived from the PIAAC background questionnaire: "Participated in formal AET in 12 months preceding survey" and "Participated in formal or non-formal AET in 12 months preceding survey". In our sample of 8,535 full-time professionals in the 14 selected countries, 1,322 reported participated in formal AET and 1,521 reported did not participate in formal or non-formal AET. Logically, the rest of respondents should have participated in non-formal AET only. It is possible that some formal AET participants also participated in non-formal AET. However, we can not distinguish this particular group from formal AET participants based on the information available at the time of this study. Moreover, due the relatively few formal AET participants in the sample, we want to over-sample formal AET participants to address the class imbalance problem. Therefore, it makes sense to consider those participated both formal and non-formal AET as formal AET participants in this specific study.

RQ1c. How do the relationships between full-time professionals' PS-TRE proficiency scores and their proficiency scores in Literacy and Numeracy vary across the 14 selected countries?

RQ1d. How do the distributions of full-time professionals' PS-TRE proficiency scores compare to those of full-time associates in the aforementioned occupational sub-categories (except Teaching)?

RQ2. For full-time professionals within each country, how do their socio-demographic background factors, occupational categories, and the types and intensities of skills use at work (and at home), relate to their probabilities of participating in the different formats of AET programs?

RQ3. Comparing full-time professionals estimated to have similar probabilities (i.e., propensity scores) of participating in a certain format of AET programs, do participants of certain format of AET programs score higher than non-participants as measured by PIAAC PS-TRE?

(Note: In RQ4, parallel analyses will be conducted with (1) full-time associate professionals' PS-TRE proficiency scores and (2) full-time professionals' Literacy and Numeracy scores as the outcome)

RQ4. To what extent are the estimated associations between full-time professionals' PS-TRE proficiency scores and their participation in the different formats of AET programs robust to hidden bias due to unobserved confounders?

Because propensity score methods only remove bias due to observed confounders and the components of unobserved confounders that are correlated with observed one, it is important to determine how strong the effect of a hypothetically omitted confounder would have to be to produce a "material change"[13] in the conclusions of the study (Rosenbaum, 2010; Rosenbaum & Rubin, 1983). In this study, the robustness of the estimated associations between full-time professionals' PS-TRE proficiency scores and their participation in the different formats of AET programs is checked via parallel analyses with (1) a similar but different sample (i.e., full-time associate professionals in the 14 selected countries) and (2) two alternative outcomes that are supposed to be highly correlated with PS-TRE proficiency Scores (i.e., Literacy and Numeracy proficiency scores as measured by PIAAC).

1.5 Significance of the Study

This study is distinct from previous research on large-scale assessment surveys of adult skills in at least four ways:

1. It draws upon data from PIAAC which, for the first time, offered a direct measure of problem solving in technology-rich environments (PS-TRE). In the face of human capital challenges of the 21st century, discussions of adult skills and lifelong learning have increased significantly in number and sophistication at the local, national and global levels. However, a review of the current literature found that a preponderance of studies focused on traditional cognitive skills. This study will fill a gap in the literature and provide a new perspective on the more comprehensive, yet comparatively under-investigated realm of problem-solving in technology-rich environments.

[13] In this study, the "material change" is defined as changing the statistical significance of the finding and/or the magnitude of the estimated association substantively and meaningfully.

2. It will only focus on full-time workers in one of the six occupational sub-categories of "Professionals". The sample homogeneity makes it possible to extend the scope of the investigation beyond respondents' socio-demographic background and focus on their work-related experiences, including skills use at work and AET participation. The significant results, if there are any, will suggest strategies for policy-makers and employers to evaluate full-time professionals' skills proficiency using a more robust and actionable metric, rather than relying solely on educational attainment.

3. By virtue of the job-requirement approach (JRA) adopted by PIAAC, it is also the first time that the types and intensities of skills use at work will be examined in great detail. The broad coverage -- from key information-processing to generic workplace skills -- provides more objective information regarding the task content of the main job held by the respondent (OECD, 2013a).

4. Last but not least, this study takes into consideration the diversity in adult learning activities and makes an effort to distinguish among the different formats of AET programs. Advanced statistical techniques (i.e., propensity score methods) will be employed to reduce bias in the estimation of the associations between the different formats of AET participation and full-time professionals' PS-TRE proficiency. Hopefully, the results of this study will shed some light on the content and organization of AET programs that aim to optimize the skills maintenance and development of full-time professionals.

From the perspectives of policy-makers and stakeholders in education and workforce development, this study has the potential to contribute to (1) a better understanding of current job tasks composition and best work practices across professional fields, both within and cross countries and (2) a more thorough understanding of the role that the different formats of AET programs played in maintaining and developing full-time professionals' problem solving skills in technology-rich environments.

1.6 Limitations of the Study

It is clear that the study faces clear limitations and challenges. Perhaps the biggest limitation is that the cross-sectional nature of the PIAAC data makes it "impossible to infer strict causality in any observed relationships between variables" (OECD, 2009). Due to the lack of random assignment, we cannot necessarily rule out the possibility of reverse causation and that the factors that led to differences in group membership may also account for differences in the outcome (i.e., confounders). Although several approaches have been developed to address this limitation (e.g., propensity score methods), making causal inferences from cross-sectional data still present a challenge to drawing policy-relevant conclusions from observational studies. What is more, the existence of unobserved, unmeasured, or unmeasurable confounding variables further constrains causal inferences from observational studies. Even the most sophisticated statistical techniques cannot fully adjust away significant confounding.

Second, some information collected via the PIAAC survey may be subject to recall bias; for instance, such as parental education and the use of skills at home. These self-reported data should be considered in light of some methodological limitations. And it is critical to always be aware of these biases and limitations in interpreting the results. With rigorous research methodologies (including study design and data collection), relevant inferences can still be drawn from cross-sectional data sets, but must be accompanied by appropriate measures of uncertainty.

1.7 Dissertation Organization (Roadmap)

Chapter 1 starts with a brief review of previous attempts to define problem solving, followed by the introduction of PIAAC and its definition of problem solving in technology-rich environments (PS-TRE). The chapter then lists the subjects of interest, research objectives and research questions, as well as the significance and limitations of this study. Chapter 1 ends with an outline of the organization of the dissertation (roadmap).

Chapter 2 lends support from the literature to the research design and methods described in the subsequent chapter. It has a dual focus on content knowledge and methodological rigor. The first section provides an in-depth examination of the theoretical framework for factors related to PS-TRE. The chapter then introduces observational studies as an alternative to randomized controlled trials, followed by a discussion of the challenges in drawing causal inferences from large observational studies. The last two sections elaborate statistical techniques tailored to address bias, both overt and hidden, which threatens the validity of causal inferences due to the observational nature of the PIAAC data.

Chapter 3 features a detailed exposition of research design and methods to address the research questions of interest. The first two sections briefly define the sample of full-time professionals from the 14 selected countries and introduce the five blocks of factors that will be used in the modeling process. The chapter then explicates the corresponding statistical techniques for each of the four research questions. The rest of Chapter 3 is devoted to a preliminary analysis examining descriptive statistics of the five blocks of factors for the sample of full-time professionals in the 14 selected countries. To facilitate parallel analyses with full-time associate professionals in research question three, descriptive statistics of the five blocks of factors for the associated sample is also provided at the end of this chapter.

Chapter 4 presents empirical results for each of the four research questions. Each research question is answered via a series of statistical analyses and corresponding tables and figures are presented to aid the understanding of the results. Finally, Chapter 5 concludes this study with a summary of the results, discusses limitations of current studies and suggest potential directions for future research.

Chapter 2. Literature Review

2.1 Theoretical Framework for Factors Related to Problem Solving in Technology-Rich Environments (PS-TRE)

Given the economic and social significance of problem solving skills in today's technology-intensive labor markets, extensive research has been devoted to identifying factors related to problem solving skills. With the study research questions in mind, Chapter 2 will start by describing the theoretical framework that provides the underpinning for variable selection and model building in Chapter 3.

An overview of previous studies on skill acquisition, maintenance and decline summarized the underlying factors into five main areas:

- Socio-Demographic Background Factors, including age, gender, family background (proxied by parental education) and educational attainment.
- Occupational Categories classified under "Professionals" by the ISCO-08 framework.
- Use of Key Information-Processing Skills at Work (and at Home), including reading and writing, numeracy and ICT.
- Use of Generic Workplace Skills/Job-related Activities, including influencing, planning, task discretion and learning at work.
- Participation in formal and non-formal AET programs.

2.1.1 Socio-Demographic Background Factors

Age (in 10-year bands):

It is universally acknowledged that age should be a basic demographic variable in the analysis of adult skills survey. The *First Results from the Survey of Adult Skills* (OECD, 2013c) showed that there are significant age group vs. observed score-point differences in key information-processing skills between the young and the old cohorts. For example, on average across the 23 participating countries[14], older adults score lower on the literacy scale than younger age groups. The average score among 55-65 year-olds is 255 points (Level 2), among 45-54 year-old is 268 points (Level 2). By contrast, the average scores for adults aged 35-44 (279 points), 25-34 (284 points) and 16-24 (280 points) all correspond to Level 3 (OECD, 2013c). Such age-related proficiency gaps are particularly marked in the domain of PS-TRE. On average across the 19 countries that participated in the PIAAC problem-solving assessment, 51% of young adults (16-24 year-olds) score at Level 2 or higher on the PS-TRE scale. By contrast, 12% of older adults (55-65 year-olds) score at Level 2 or higher. Note that Level 2 on the proficiency scale does not represent a performance benchmark, the division between Lever 2 and below and Level 3 and above in Literacy and Numeracy and Level 2 and above and Level 1 and below in PS-TRE has been made for ease of presentation (OECD, 2013c). Please refer to Chapter 21 of the *Technical Report of the Survey of Adult Skills (PIAAC)* (OECD, 2013a) for a detailed description of the content definition for each proficiency level per cognitive domain.

---

[14] Due to sampling issues, the Russian Federation is not included in the cross-country calculations presented in any of the figures or tables in the *First Results from the Survey of Adult Skills* (OECD, 2013c). Detailed information can be found in the *Technical Report of the Survey of Adult Skills* (OECD, 2013a).

The OECD (2013c) also found that the size of the gap in proficiency levels between the younger and the older cohorts "change(s) little and remain(s) substantial" when other socio-demographic background factors, such as educational attainment, are taken into account. While it is plausible to attribute much of these strong age-related differences in proficiency to biological aging, this is surely more to this phenomenon. Although older adults generally have lower proficiencies than their younger counterparts, there are wide variations in the extent of gap across countries. For instance, 16-24 year-olds in Korea have the greatest advantage in proficiency over 55-65 year-olds – 49 points higher on both the literacy and numeracy scales, with 60% more scoring at Level 2 or higher on the PS-TRE scale. By contrast, England/Northern Ireland (UK) shows no significant differences by age cohort on the literacy and numeracy scales, while the United States shows the smallest difference (i.e., 17%) in the percentage of adults scoring at Level 2 or higher in PS-TRE between the two groups. (OECD, 2013c).

The cross-country variations in proficiency gaps between the younger and older cohorts suggest the existence of other potential factors that are associated with both age cohorts and skills proficiency (i.e., cohort effect). One explanation is that changes in the amount of formal education received by different age groups may vary in each country. For example, Korea, the country with the largest generation gap in proficiency, has been known for its success in raising the educational attainment of the younger cohorts. In 1970, only 6% of its labor force had a university-level education. By 2010, 65% of 25-34 year-olds in Korea have completed tertiary education (OECD, 2013c). In addition to the quantity of education received by different age cohorts, different degrees of improvements in the quality of initial education over time may also account for age-related variations in skills proficiency in different countries. What is more, depending on the adult education system of each country, learning opportunities offered to adults "to undertake further training or to engage in practices that help to maintain and develop proficiency over their lifetimes" (OECD, 2013c) may vary in both quantity and quality.

The performance of adults grouped by age cohorts is comparable to performance of their counterparts in other international large-scale assessments. The *Final Report of the International Adult Literacy Survey* (OECD & Statistics Canada, 2000) showed that in every participating country of the IALS when only age is considered, younger adults aged 26-35 have higher literacy scores than adults closer to retirement aged 56-65. The document (OECD & Statistics Canada, 2000) also noticed there are significant differences across countries: Belgium, Canada, Finland, Poland and Slovenia have the largest differences between the mean literacy scores for the two age groups (greater than 50 points) while New Zealand and the United States have the smallest difference (less than 20 points).

Like in the Survey of Adult skills (a product of PIAAC), the observed differences between age groups could be a mixture of ageing and cohort effect. Since cohort effect cannot be distinguished in a cross-sectional survey, Paccagnella (2016) came up with synthetic cohorts comprise countries that have participated in two surveys -- the IALS (administered 1994-1998) and the PIAAC (administered in 2012). Note that comparing results from the IALS and the PIAAC is possible "because the literacy scales share items and can be put on the same scale (OECD, 2013b). It is found that older adults in PIAAC are generally more proficient than their age group counterparts in IALS. The author attributed such differences to cohort effect partially due to "the increase in educational attainments that took place over recent decades". However, a certain degree of caution should always be exercised in interpreting these results due to PIAAC's switch of delivery mode from paper-based to computer-based.

Gender

Gender is another important factor to consider when analyzing data from the assessment of competencies. Results from the first PIAAC survey show that on average, men score marginally higher than women on the literacy scale by 2 points. The nearly negligible gender gap is the result of significant social progress made over the past few decades. (OECD, 2013c). However, gender-related proficiency gaps persist in the domains of numeracy and problem solving. On average across the 23 participating countries, men score 13 points higher on the numeracy scale than women. On average across the 19 countries that participated in the PIAAC problem-solving assessment, 36% of men are proficient at Level 2 or higher on the PS-TRE scale, compared to 32% of women (OECD, 2013c).

Given decades of efforts made to close the gender gap in academic achievement, it is not surprising to find that gender-related differences in skills proficiency tend to be smaller in the youngest cohort than in the entire population surveyed (OECD, 2013c). In the domain of literacy particularly, the gender gap in favor of men virtually disappeared among young adults; "and where there are small differences, it is young women who have higher scores". In the domains of numeracy and problem solving, on the other hand, men still have better results than women, even for the youngest age cohort. The OECD document (2013c) suggested that greater computer use among men probably contributes to gender differences in proficiency in PS-TRE. The plausibility of this explanation will be examined in further detail in Chapter 3.

Similar patterns of gender disparities in skills proficiency have been seen in other national and international large-scale assessments. Data from the 2012 National Assessment of Educational Progress (NAEP) showed no significant differences in mathematics scores at ages 9 or 14 in the U.S. However, a significant gender gap in mathematics achievement was found at age 17, with male students scoring higher than female students (National Center for Education Statistics, 2013). Another study (Arora & Pawlowksi, 2017) combining data from PISA and PIAAC found that gender differences in mathematics performance of the 15-year-olds cohort in PISA 2003 “either stayed the same in PIAAC 2012 (when those in the cohort were 23 to 25 years old) or increased”. They also found that the size of the gender gap in numeracy increases as age increases, with the 16-24 age group showing the least differences within the total PIAAC population. All these findings suggest that there is still a long way to go toward gender equity in skills proficiency.

Family Background (proxied by Parental Education)

The socio-economic background in which adults were raised has a profound influence on the development of their skills over a lifetime. Findings on the association between socio-economic background and academic performance abound in the literatures of education, sociology and psychology. In examining America’s skills gap, Kirsch et al. (2016) looked at various subpopulations of students and found “persistent gaps in academic achievement across … socioeconomic groups”. In fact, neighborhoods have played such a critical role in American students’ transmission of opportunity that prompted the authors to caution “segregation by socioeconomic status has been on the rise”. The positive effects of exposure to better neighborhood on shaping the socioeconomic status of children as they grow to adulthood are documented in many studies. See also the evidence from the Moving to Opportunity Experiment conducted by the National Bureau of Economic Research (Chetty, Hendren & Katz, 2015).

In the Survey of Adult Skills (aka PIAAC), it is clear that “adults from socio-economically advantaged backgrounds have higher scores in key information-processing skills on average than those from disadvantaged backgrounds". (OECD, 2013c). Like most large-scale surveys, the PIAAC uses parental education as a proxy for family background. The rationale is that growing up in a family with highly educated parents offers opportunities to achieve one's potential, including proficiency in key information-processing skills (OECD, 2013c). On average across the 23 participating countries, adults with at least one parent who had attained tertiary education score 40 points higher in literacy than those with neither parent having attained upper secondary education. An average of 16% of adults with neither parent having attained upper secondary are proficient at Level 2 or higher on the PS-TRE scale, in contrast to 55% of adults with at least a tertiary-educated parent score at this level or higher. In a more recent study, Braun (2018) examined the relationship between parental education and skill levels within each educational attainment stratum and found "the association is stronger at lower levels of Educational Attainment and more pronounced in the contrast between the lowest and highest levels of Parental Education".

It is important to note that the so-called socio-economic gradient (i.e., the association between skills proficiency and socio-economic background) varies widely across countries and age groups. On average across countries, the slope "is steeper for the adult population as a whole than for young people" (OECD, 2013c). However, the picture is mixed at the individual country level. The United States, for example, has the steepest gradient (indicating the strongest relationship) among 16-65 year-olds but is close to the OECD average among 16-24 year-olds. The Czech Republic, Denmark, England/Northern Ireland (UK), Estonia and the Slovak Republic, by contrast, have a steeper socio-economic gradient among young people than among the full adult population (OECD, 2013c). In general, a flatter socio-economic gradient among younger cohorts signals greater socio-economic mobility among younger cohorts.

Educational Attainment

Last but not least, the role of educational attainment in fostering proficiency in key information-processing skills cannot be overestimated. Formal education "strengthens information-processing skills directly, through the coursework involved", and indirectly, through "better jobs with possibly more opportunities to develop these skills" (OECD, 2013c). In fact, the indirect relationship between educational attainment and skills proficiency is phrased as the "cumulative advantage in key information-processing skills for high-educated adults" when the combination of higher education levels and better opportunities to improve proficiency has the potential to evolve into a virtuous circle, in which higher proficiency leads to more opportunities to further develop proficiency (OECD, 2013c).

On average across countries, adults who have attained tertiary education (i.e. highly educated) score 25 points higher in literacy than upper secondary graduates and 51 points higher than those who have not attained upper secondary education (i.e., low-educated). Even after adjusting for other socio-demographic background factors (e.g., age, gender, family background/parental education, and occupational category), adults with tertiary-level education retain a 36-point advantage on the literacy scale over those who had not attained upper secondary education. To put the adjusted differences in perspective, "a 36 score-point difference is estimated to be the equivalent of around five years of additional education" (OECD, 2013c). On the scale of PS-TRE, 52% of tertiary-educated adults score at Level 2 or higher compared to only 19% of adults who have not attained upper secondary education score at this level or higher (OECD, 2013c).

The relationship between educational attainment and proficiency in key information-processing skills is undoubtedly positive and strong; however, the strength of the relationship varies considerably among countries. Possible explanations include cross-national "differences in the quality of schooling, the nature of adult-learning systems, and differences in patterns of participation in education" to name a few (OECD, 2013c). More interestingly, the proficiencies of adults at the same level of educational attainment (and similar socio-demographic background factors) may vary substantially among countries. In addition to attributing the variations to the aforementioned country-level differences, it is advisable to also take into consideration occupation-level differences that will be explored in the next section.

2.1.2 Occupational Categories Classified under "Professionals" by the ISCO-08 Framework

As is the case with educational attainment, the relationship between type of occupation and proficiency in key information-processing skills is two-fold. On the one hand, skills proficiency, to a great extent, determines the type of job one can get. On the other hand, adults holding jobs requiring higher skills proficiency tend to have "more opportunities for using, thus maintaining and developing, literacy, numeracy and problem-solving skills" (OECD, 2013c). The aforementioned cumulative advantage in key information-processing skills for high-educated adults provides a partial explanation for the complex relationship. As mentioned before, higher proficiency in key information-processing skills is closely related to higher educational attainment which, in turn, is closely related to adults' possibility of securing employment in high-skilled occupations.

Occupation-related differences in skills proficiency are universal across countries and age groups. The *First Results from the Survey of Adult Skills* (OECD, 2013c) distinguished among four types of occupation (skilled, semi-skilled white-collar, semi-skilled blue-collar and elementary occupations) and found that in all countries, adults in skilled occupations score higher, on average, than those in semi-skilled or elementary occupations, in both literacy and numeracy. On average across countries, 50% of adults in skilled occupations score at Level 2 or higher, while 20% of those in elementary occupations score at this level or higher.

As mentioned in Section 1.3, this study will mainly focus on “full-time professionals” -- workers of participating countries who reported their current status as “full-time employed (self-employed or employee)”. This narrowed focus allows us to conduct a fine-grained examination of occupational categories classified as “Professionals” by the ISCO-08 framework. A list of the six sub-categories under the heading of “Professionals” will be displayed in Table 4. As expected, all full-time professionals in the sample work in skilled occupations.

In spite of the homogeneity of skill levels required to undertake these professional jobs, individuals may vary considerably in subject-specific professional skills. For instance, it is expected that professionals in the information and communication technology industry will demonstrate higher proficiencies in solving problems in technology-rich environments; while legal, social and cultural professionals will have an advantage in reading and writing. Controlling for occupational category allows a closer look at distinctive patterns of skills use at work that are associated with PS-TRE proficiency scores, without the confounding effect of full-time professionals’ skill specialization.

2.1.3 Use of Key Information-Processing Skills at Work (and at Home)

In addition to contextual factors such as socio-demographic background and occupational categories, productive use of skills at work (and at home) has been proven to be a key “both for developing proficiency and preventing its loss” (OECD, 2013c). It is no secret that with the advance of technology, the character of “in-demand skills” is constantly evolving. The dynamics of the market economy often result in a time lag or mismatch between employees' "own skills" and "job skills" that are actually used in the workplace (OECD, 2013a). Such demand-supply mismatch is especially common among young adults whose supply of skills comes mostly from formal education. While the level and quality of formal education is crucial in laying the cornerstone for adults' skills proficiency, there is still room for growth by putting skills into productive use at work (and at home). While acknowledging the "close positive relationship" between educational attainment and proficiency in information-processing skills, the *First Results from the Survey of Adult Skills* (2013c) warned that "educational qualifications do not necessarily reflect the level of an individual's literacy, numeracy or problem-solving skills (p.199) and "skills developed in formal education can depreciate if they are not used" (p.118). Chapter 5 of the book discussed the direct and indirect role of education in fostering information-processing skills in more detail. By virtue of the task-based JRA, the Survey of Adult Skills (PIAAC) measures the use of a wide range of skills at work, ranging from key information-processing skills to generic workplace skills. To condense the large amount of information collected through the JRA module, eight indicators of skills use at work – four for key information processing skills and the other four for generic workplace skills -- were constructed that "group together tasks associated with the use of similar skills" (OECD, 2013c). Table 2 lists the eight skill-use indicators at work along with sample tasks from which these indices were derived.

Table 2. Overview of Indicators of Skills Use at Work

	Indicator	Group of Tasks
Key Information-Processing Skills	**Reading**	Reading documents (directions, instructions, letters, memos, e-mails, articles, books, manuals, bills, invoices, diagrams, maps)
	Writing	Writing documents (letters, memos, e-mails, articles, reports, forms)
	Numeracy	Calculating prices, costs or budgets; use of fractions, decimals or percentages; use of calculators; preparing graphs or tables; algebra or formulas; use of advanced math or statistics (calculus, trigonometry, regressions)
	ICT	Using e-mail, Internet, spreadsheets, word processors, programming languages; conducting transactions on line; participating in online discussions (conferences, chats)
Generic Workplace Skills/Job-related Activities	**Influencing**	Instructing, teaching or training people; making speeches or presentations; selling products or services; advising people; planning others' activities; persuading or influencing others; negotiating
	Task Discretion	Choosing or changing the sequence of job tasks, the speed of work, working hours; choosing how to do the job
	Learning at Work	Learning new things from supervisors or co-workers; learning-by-doing; keeping up-to-date with new products or services

The *First Results from the Survey of Adult Skills* (OECD, 2013c) confirmed that "adults who engage more often in literacy- and numeracy-related activities and use ICTs more have higher proficiency in literacy and numeracy". On the Literacy scale particularly, adults who frequently use ICT skills at work scored 15 points higher (on average across countries) than those who never do, when other socio-demographic background and practice-oriented factors are taken into account. In the belief that a strong relationship should also be found concerning proficiency in problem solving in technology-rich environments, this study will further analyze results on the PS-TRE scale and observe how the strength of the relationship varies across countries.

Similar to educational attainment, there appears to be a cumulative advantage in frequent practice of reading, writing, numeracy and ICTs at work. On one hand, intensive use of key information-processing skills should aid in maintaining and developing proficiency in these skills; on the other, having higher levels of proficiency in key information-processing skills is likely to result in greater opportunities to engage in such work activities more frequently. This reciprocal relationship could be a source of concern in that it creates a vicious circle where low proficiency is reinforced by limited use at work with fewer opportunities for improvement (OECD, 2013c).

2.1.4 Use of Generic Workplace Skills/Job-related Activities

The mutually reinforcing relationship between the use of, and proficiency in, key information-processing skills supports Reder's point that the best way to maintain and develop skills is to use them (2009a; 2009b). Unlike key information-processing skills which are measured directly, generic workplace skills/job-related activities are only self-reported and documented in the JRA module. Moreover, while key information-processing skills "tend to be used together", the use of generic workplace skills/job-related activities are found to vary considerably across skill domains (OECD, 2013c). Such heterogeneity is informative because it reflects job-specific skills distinctively required by different categories of occupation, over and above key information-processing skills which are foundational in all jobs.

As shown in OECD (2013c, Figure 4.18), the use of influencing, planning, task discretion and learning at work skills increases substantially from elementary occupations to professional occupations. However, the broad occupational categories fall short of capturing differences across the six sub-categories of "Professionals". This study will fill the knowledge gap by conducting more nuanced comparisons of the types and intensities of generic workplace skills/job-related activities across the six occupational sub-categories of “Professionals”.

2.1.5 Formal and Non-formal AET Participation

Over the last two decades, drastic structural changes have swept through the global labor markets and resulted in a strong push towards lifelong learning. In order to compete in increasingly knowledge-based economies, many countries and organizations have invested heavily in adult education and training (AET) programs designed to upskill and/or reskill their workforce according to the needs of the jobs for now and for the future. According to the definition of UNESCO (2011, 2012), there are two major forms of education and training -- formal education and training comprises education that is institutionalized, intentional and planned through public organizations and recognized private bodies; while non-formal education is institutionalized, intentional and planned by an education provider. Non-formal education mostly leads to qualifications that are not recognized as formal qualifications by the relevant national educational authorities or to no qualifications at all.

Despite a growing emphasis on the necessity of participating in continuous learning activities, empirical evidence on the relationships between AET programs and skills proficiency is scarce. In fact, there exist some doubts concerning returns from “costly and hardly reversible investment” in AET programs compared to investment in school and pre-school education (Torgerson et al., 2004). To address these concerns, a few recent studies have drawn upon the PIAAC data for their cross-national comparative research. Referring to adult education and training experienced by respondents in the 12 months preceding the survey, the *First Results from the Survey of Adult Skills* (OECD, 2013c) showed “a strong positive relationship, consistent across countries”, between AET participation and literacy proficiency. Later on, Sgobbi (2014) narrowed her investigation to eleven EU countries and found that formal AET has "either negative or non-significant impact on cognitive skills" across countries, where non-formal AET displays "significant and positive effects."

This finding was echoed in Cegolon's paper (2016) where only four of the eleven EU countries were investigated. Using OLS and quantile regressions which consider a set of covariates including age, gender and years of schooling, the research find that formal AET has "a smaller impact" on both literacy and numeracy compared to non-formal AET. What is more, the results showed that both types of AET programs "take different trajectories in each of the countries selected". In Italy and France, AET programs seem to be more efficient for adults "at the top of the skill distributions"; whereas in Sweden and the UK, "the differences are less marked across all distributions, suggesting a fairer effect of both types of AET".

While all three studies support they hypothesis that non-formal AET is a significant driver of individual proficiency in cognitive skills, Sgobbi (2014) noted that in comparison to literacy and numeracy, "performance in problem solving is less dependent on achievements and opportunities of early age and more easily developed through experience, adult education and training initiatives". These findings have two important implications:

(1) Although problem solving is not a domain typically taught as a school subject, it is a comprehensive skill that can be developed and improved continuously through work-related experiences;

(2) AET may provide a remedy for educational inequity resulting from a disadvantaged background in early life (Cegolon, 2016).

Last but not least, although there is a large literature suggesting that AET participation is associated with high level of cognitive skills, the common dependence of training opportunities and skills proficiency on employment is often omitted. To account for possible systematic differences between individuals of a certain job status and the remaining adult population, it is meaningful to examine the relationship between AET participation and skills proficiency within a specific group. For example, a mixed picture emerges when restricting the analysis to employed and self-employed individuals. Sgobbi (2014) found that "the impact of adult education and training on cognitive skills does not substantially change" whereas Cegolon (2016) found that "the effect of training gets smaller and less significant when the analysis is restricted to employed individuals". Both results could be justified depending on how they control for other variables that may significantly affect the relationship. In Chapter 3, we will focus on full-time professionals because of "their significant larger exposition to training and adult education initiatives, as well as by … the more immediate economic impact expected from training this population group" (Sogbbi, 2014). Advanced statistical techniques will be discussed to deal with potential endogeneity of AET participation.

2.2 Challenges in Drawing Causal Inferences from Observational Studies

2.2.1 Observational Studies as an Alternative to Randomized Controlled Trails

The birth of the modern randomized controlled trial (RCT) in medical and social science settings can be traced back to 1948 when researchers at the British Medical Research Council (MRC) "replaced alternate allocation[15] with strict concealed randomization of patients to treatment and control groups" (Bothwell & Podolsky, 2016). The blinding of researchers to subjects' assignments due to chance (i.e., concealed randomization), if at all possible, soon

[15] Alternative allocation, the most recent methodologic ancestor of RCTs, entailed treating every other subject with a particular treatment, withholding it from the others, and then comparing outcomes. However, selection bias stemming from the ease of cheating the process remained a major limitation of alternate allocation (Bothwell & Podolsky, 2016, Bothwell et al., 2016).

emerged as a solution to reduce spurious causality and bias (Podolsky et al., 2016). By the late 20th century, RCTs were recognized as the gold standard for clinical trials and have increasingly appeared in the literature of agriculture, due to Fisher's (1925) experimental research and his writings that popularized randomized experiments.

Notwithstanding the wide acceptance of RCTs as the gold standard for internal validity across disciplines, the past seven decades also bear witness to many limitations of RCTs (Ginsburg & Smith, 2016). Frist of all, it is not always practical or ethical to randomly assign participants to either a treatment or control group in social settings. For example, one of the most important issues raised by RCTs involving human subjects is whether it is ethical to withhold effective treatments from research subjects in order to satisfy scientific objectives (Resnik, 2008). Second, RCTs can be demanding in terms of the financial and human resources needed for research design, record keeping and ethical review that minimizes risks and protects privacy and confidentiality (Bothwell et al., 2016). As a result, RCT disproportionately reflect the interests of industrialized regions (Petryna, 2009). Third, even well-conducted RCTs sometimes failed to influence practice because of the discrepancy between the time frame of RCTs and the fast pace of innovation (Bothwell et al., 2016). Last but not least, the strict eligibility criteria of RCTs create a highly controlled setting/artificial environment which may not always mimic real life situation. Consequently, RCTs can return results that are seldom generalizable to the general population (i.e., limited external validity) (Clay, 2010).

RCTs have significant limitations when conducted in more realistic settings and these limitations have prompted researchers to pursue alternative ways to optimizing the trade-off between internal validity (the ability to trace causal inferences to the intervention) and external validity (the generalizability of the results) (Clay, 2010). Among them, large nonrandomized observational studies are increasingly used in social and behavioral sciences and medicine as an alternative to RCTs. Compared to RCTs, observational studies are relatively more efficient and less expensive to perform. More importantly, the literature has suggested a high degree of correlation between treatment effects reported in observational studies and RCTs. Ioannidis et al. (2001) found significant correlation (Spearman coefficient of .75, $p < .001$) between treatment effects reported in RCTs versus observational studies across 45 diverse topics in general internal medicine. Ioannidis et al.'s results were echoed by a more recent Cochrane meta-analysis (Anglemyer, Horvath & Bero, 2014) that found no significant difference in effect estimates between RCTs and observational studies regardless of the observational study design.

2.2.2 Challenges in Drawing Causal Inferences from Large Observational Studies

Favorable comparisons of observational study results to RCT findings have created novel opportunities for researchers and policy makers. The utility of large observational data in social science research, in particular, has permitted inquiries into research questions otherwise untestable under the RCT framework and, therefore, expands the scope of scientific inquiry. However, statistical analysis of such data is not without its challenges. The lack of random assignment in observational studies makes it very difficult to draw credible conclusions about causal connections.

According to Rosenbaum (2002), an observational study is an empirical investigation of the effects caused by a treatment, policy, or intervention in which it is not possible to assign subjects at random to treatment or control, as would be done in a controlled experiment. The term *observational study* includes two types of research designs. Studies where (1) subjects are not randomly assigned to conditions and (2) the independent variable is manipulated before the dependent variable is measured are referred to as quasi-experimental designs (Cook & Campbell, 1979). On the other hand, studies that lack both random assignment and manipulation of conditions are referred to as non-experimental design.[16] Observational studies are common in most fields that study the effects of treatment on people; however, there are a few methodological issues inherent in conducting research with observational data:

- Selection bias, can occur if subjects are not assigned to treatment or control at random. Differences in the outcome may be due, in part, to pre-treatment differences rather than treatment effects (Rosenbaum, 2005). For example, since AET participation is not randomly assigned to individuals but instead is the consequence of a variety of socio-demographic background and work-related factors, it is difficult to know definitively whether differences in AET participation are responsible for differences in skills proficiency. Note that selection bias is distinct from confounding bias in that the former compromises external validity while the latter compromised internal validity (Haneuse, 2016).
- Confounding bias, or the failure to appropriately account for the presence of additional variables that are associated with both treatment choice and the outcome of interest. Confounding bias undermines the accuracy of the model estimates (i.e., internal validity). In this study, we are interested in the association between AET participation and full-time professionals' skills proficiency. On the one hand, there is substantial evidence that individuals with higher educational attainments are more likely to participate in AET programs (Bassanini et al., 2007; Dieckhoff & Steiber 2011; Albert et al., 2010). On the other, it has been

[16] In this paper, the terms observational study, quasi-experimental design and non-experimental design will be used equivalently.

shown that “educational attainment has a strong positive relationship to proficiency … after other characteristics have been taken into account” (OECD, 2013c). As a result, AET participation will appear to be associated with a higher probability of achieving skills proficiency. Note that such results are very likely not due to the independent effect of AET programs but rather the combined effects of educational attainment and AET participation. The indirect relationship between educational attainment and skills proficiency, through AET participation, is termed as the “cumulative advantage” in Section 2.1.

2.3 Propensity Score Methods to Reduce Bias

Because subjects are not randomly assigned to conditions, estimates of treatment effects may be biased due to “nonrandom differences between treated and untreated groups with respect to covariates related to the outcome” (Leite, 2017). To reduce bias in observational studies, Rosenbaum and Rubin (1983; 1984; 1985) proposed the use of propensity score which, essentially, is the estimated probability of treatment as a function of observed confounders. Over the last three decades, propensity score methods have become a common choice for estimating treatment effects with observational data in social science research (Thoemmes & Kim, 2011). In this section, we will introduce the common theoretical foundation for all variations of propensity score methods that will be described in Chapter 3

2.3.1 Rubin’s Causal Model

Propensity score methods are based on Rubin’s earlier work (1973a; 1973b; 1974) concerned with problems of causal inference. Many empirical studies in social sciences are interested in evaluating the effects of a treatment, policy, or intervention on some pre-defined outcomes. In an ideal situation, each subject can be exposed to one or more different levels of the treatment (Imbens & Wooldridge, 2009). Take AET participation for example: An individual might choose to or choose not to participate in an AET program. The object of interest will be a comparison of the two potential outcomes (i.e. PS-TRE proficiency scores) for the same individual when she participated, and when she didn't participate in the AET program. In reality, however, it is often impossible to have an individual who is both a participant (i.e., in the treatment group) and a non-participant (i.e., in the control group). This problem is referred to as the "fundamental problem of causal inference" (Holland, 1986).

In a series of papers, Rubin (1973a; 1973b; 1974) proposed the potential outcomes (or counterfactual) framework also known as Rubin's Causal Model (Holland, 1986; Shadish, 2010). In Rubin's Causal Model, all individuals in the population have potential outcomes associated with the different levels of exposure to the treatment. Treatment effects are then interpreted as the differences between pairs of potential outcomes for the same individual given the presence and absence of treatments. In the simplest form, each individual $\boldsymbol{i}$ has two potential outcomes: $\boldsymbol{Y_{1i}}$ for the treatment condition and $\boldsymbol{Y_{0i}}$ for the control condition. Therefore, the treatment effect for this individual is $\boldsymbol{Y_{1i} - Y_{0i}}$. Note that since each individual can only be in one condition at any moment in time, the unobserved potential outcome is hypothetical or counterfactual. Table 3 from Morgan and Winship (2007) illustrated the basic concept of observed and counterfactual outcomes.

Table 3. Potential Outcomes in the Presence and Absence of a Treatment

Group	Y_{1i}	Y_{0i}
Treatment (T_i=1)	Observed	Counterfactual
Control (T_i=0)	Counterfactual	Observed

Based on the framework of Rubin's Causal Model, the observed outcome for individual *i,* Y_i, is the observed value in either the treatment condition or the control condition:

$$Y_i = T_i Y_{1i} + (1-T_i) Y_{0i}$$ **Equation 1**

where T_i indicates whether individual *i* is in the treatment group ($T_i = 1$) or the control group ($T_i = 0$).

On the other hand, the expected values of the (unobserved) counterfactual outcomes, $E[Y_{0i} \mid T_i = 1]$ and $E[Y_{1i} \mid T_i = 0]$, are estimated using the expected values of the observed outcomes for each group, $E[Y_{0i} \mid T_i = 0]$ and $E[Y_{1i} \mid T_i = 1]$, respectively. The average treatment effect (ATE) is then defined as the difference between the expected values of the potential outcomes of all individuals in the treatment group and all individuals in the control group:

$$ATE = E[Y_{1i}] - E[Y_{0i}]$$ **Equation 2**

The average treatment effect on the treatment group (ATT) is defined as the difference between the expected values of the potential outcomes of treated individuals:

$$ATT = E[Y_{1i} \mid T_i = 1] - E[Y_{0i} \mid T_i = 1]$$ **Equation 3**

The average treatment effect on the control group (ATC) is defined as the difference between the expected values of the potential outcomes of the untreated individuals:

$$ATC = E[Y_{1i} \mid T_i = 0] - E[Y_{0i} \mid T_i = 0]$$ **Equation 4**

Note that ATE is a population-level comparison of mean outcomes, i.e., the difference in mean outcomes had the entire population been assigned to one treatment and had the entire population been assigned to another. The choice of ATE (over ATT or ATC) is appropriate to this specific study because we are interested in the difference in PS-TRE scores for all individuals, including formal, non-formal and non-participants of AET programs. Refer to McCaffrey et al. (2014) for more details on the differences between ATE, ATT and ATC. Morgan and Winship (2007) referred to Equation 2 as a naïve estimate of the average treatment effect in that it requires the fulfillment of three assumptions to obtain unbiased estimates of treatment effects for the population:

Assumption 1. Strong Ignorability of Treatment Assignment:

Rosenbaum and Rubin (1983) defined treatment assignment to be strongly ignorable if the following two conditions hold:

(1) The treatment assignment is independent of potential outcomes, given observed covariates X. This condition also assumes that all variables that affect the treatment assignment and potential outcomes have been measured (i.e., no omitted confounders). Note that the strong ignorability of treatment assignment assumption is only strictly met if there are no unobserved confounders.

(2) For every value of the covariates X, the probability of treatment assignment is neither 0 nor 1. Each individual has a nonzero probability to receive either treatment. This implies that no combination of covariate values should determine exactly which treatment an individual will receive or be excluded from receiving.

This assumption is crucial to the estimation of treatment effects. In experimental design where assignment to treatment is randomized, and thus independent of the covariates as well as potential outcomes, it is straightforward to obtain unbiased estimates of the ATE and the ATE is equal to the ATT and ATC. Nevertheless, in observational studies where "the creation of a comparison group follows a natural process that confounds group assignment with outcomes" (Guo & Fraser, 2015), the ignorable treatment assignment assumption is often violated. As a result, the estimates of treatment effects may be biased and inconsistent. In a later section, we will review one of the most common remedial approaches to relax this fundamental assumption: by rebalancing assigned conditions so that they become more akin to data generated by randomization.

In the context of multiple treatment conditions, the assumption of weak unconfoundedness is required which assumes that the assignment to each treatment condition is independent of the potential outcome of the respective condition (Imbens, 2000). Unlike strong ignorability of treatment assignment, weak unconfoundedness does not require that the assignment to one treatment condition be independent of all potential outcomes. Instead, it requires only pairwise independence of assignment to a specific condition and the potential outcome of that condition (Leite, 2017).

Assumption 2. Overlap:

Estimation of treatment effects also requires that covariate distributions must be similar in both the treatment and control groups, i.e., covariate balance. Obtaining adequate balance of covariate distributions between treated and untreated groups after the implementation of propensity score methods is evidence that strong ignorability of treatment assignment has been achieved given observed covariates X (Leite, 2017).

Assumption 3. Stable Unit Treatment Value Assumption (SUTVA):

There is a unique value for the potential outcome $\boldsymbol{Y_{ti}}$ corresponding to individual *i* and treatment t (Rosenbaum & Rubin, 1983). This one-to-one correspondence between potential outcome and treatment version has two implications: First, the distribution of potential outcomes for one individual is independent of the treatment assignment of another individual. Second, there are no unrepresented conditions of the treatment (Rubin, 1986).

2.3.2 Definition of Propensity Scores

Unlike randomized experiments, observational studies deal with complicated real-life situations that often result in violations of these assumptions. Consequently, biased and inconsistent estimates may occur. Reardon and Raudenbush (2009) conducted a set of simulation analyses to investigate the impact of violations of some of these assumptions and found that "modest violations of these assumptions degrade the quality of … estimates, but that models that explicitly account for heterogeneity of (treatment) effects are less affected by violations of the other assumptions". On the bright side, this challenge has motivated statisticians to develop more rigorous approaches to correct for violations under the condition of non-ignorable assignment.

Typical examples of these statistical approaches include three propensity score-based methods --matching, stratification and weighting -- which allow balancing of observed covariates between the treatment and control groups for more nuanced analyses. The propensity score, e, by definition, is the conditional probability of an individual ***i*** being assigned to a particular treatment Z conditional on observed covariates X (Rosenbaum & Rubin, 1983):

$$e_i = Pr(Z_i = 1|X_i) \qquad \textbf{Equation 5}$$

As a balancing score, the propensity score can be used to create "balanced" groups via controlling for pre-treatment imbalance on a set of observed covariates X (and the components of unobserved covariates that are correlated with the observed covariates). That is to say, given the same propensity score, the conditional distribution of observed covariates X (and the components of unobserved covariates that are correlated with the observed covariates) is the same between the treatment and control groups and any pre-treatment differences in the observed covariates (and the components of unobserved covariates that are correlated with the observed covariates) are approximately random (Rosenbaum & Rubin, 1983). Therefore, any post-treatment differences in the outcomes can be attributed to the effect of treatment (and the components of unobserved covariates that are not correlated with the observed covariates) (Morgan & Harding, 2006).

The balancing property of propensity scores "allows one to mimic some of the characteristics of an RCT in the context of an observational study" (Austin, 2011). When the treatment and control groups overlap enough with respect to observed covariates X, a straightforward estimation of treatment versus control effects that reflects adjustment for differences in all observed covariates (and the components of unobserved covariates that are correlated with the observed covariates) is allowed (Rubin, 1997). In other words, because groups are balanced on observed covariates X before treatment commences, the treatment effect can be estimated by contrasting outcomes between treated and untreated sets of individuals with the same propensity scores (Nicholas & Gulliford, 2008). Rosenbaum and Rubin (1983) have proved that assuming the treatment assignment is strongly ignorable given a set of observed covariates X, the mean difference between treated and untreated outcomes at a specific value of the propensity score is an unbiased estimated of the treatment effect at that value.

2.3.3 Comparison of Propensity Score Methods

Once propensity scores have been estimated – provided that all covariates related to treatment assignment are identified (i.e., no unobserved confounders) and an appropriate model is fitted to estimate propensity scores – balanced groups can be constructed by means of propensity score methods. To achieve adequate covariate balance between the treatment and control groups, an adequate area of common support is required which is an overlap in the range of propensity scores across groups. Note that no inferences of treatment effects can be made for observations outside of the area of common support. Therefore, the ATE is the average treatment effect for observations within the area of common support, not the entire sample. This section presents the description of the three propensity score methods, as well as a comparison in terms of common support and covariate balance.

- Matching

Propensity score matching consists of grouping treated and untreated observations with similar values of propensity scores. Depending on the data structure, different combinations of matching ratio and matching algorithm can be used. Matching ratios can be one-to-one, fixed ratio or variable ratio. Common matching algorithms are greedy matching, optimal matching and genetic matching. Most algorithms can match either with or without replacement.

For each treated case, greedy matching searches for the best available match among the untreated cases. Greedy matching contrasts with genetic matching and optimal matching in that it "seeks to minimize the distance between each pair but does not minimize the total distance between all matched pairs" (Austin, 2011). Greedy matching can be performed using either nearest neighbor matching or caliper[17] matching. Nearest neighbor matching simply pairs each treated case with untreated case of the smallest difference in propensity scores. Caliper matching "enforces a maximum distance within which matches are acceptable, usually in standard deviation unit" of the logit of the propensity score (Leite, 2017). For example, Rosenbaum and Rubin (1985) used a caliper of .25 standard deviations to remove at least 90% of bias. Within the caliper of each treated case, matches are found by selecting the untreated case with the closest propensity score. In practice, nearest neighbor within caliper matching strategy is widely used

[17] The propensity score 'caliper' is the maximum reasonable difference in propensity score which would allow the matching algorithm to generate a suitable match.

not only to improves the matching quality of the treated case under consideration, but also to enforce common support because “treated cases without any untreated cases within their caliper are discarded” (Leite, 2017). However, greedy matching does not guarantee matches with the minimum total distance between treated and untreated groups. In other words, greedy matching will match a treated case to the nearest untreated case even if that untreated case would better serve as a match for a subsequent treated case.

In contrast to greedy matching, optimal matching and genetic matching attempt to optimize the matching quality of the entire treated sample. Optimal matching (Rosenbaum, 1987; Austin, 2011) uses network flow theory to minimize the total distance between treated and untreated matched pairs. Optimal matching is commonly used for full matching, which attempts to match all untreated cases to a treated counterpart, “resulting in no loss of sample size as long as there is an adequate area of common support” (Leite, 2017). That being said, Gu and Rosenbaum (1993) have found that optimal matching did no better than greedy matching in producing balanced matched samples.

A major strength of genetic matching is that it searches for matches that optimize covariate balance (Diamond & Sekhon, 2013). Using a genetic algorithm, genetic matching minimizes a multivariate weighted distance on covariates between treated and untreated cases. Genetic matching can be used without including propensity scores and is particularly useful “when propensity score matching fails to achieve covariate balance, or propensity score estimation results in complete separation or quasi-complete separation of treated and untreated groups” (Allison, 2004).

- Stratification

Propensity score stratification (or sub-classification) consists of splitting the sample of treated cases and untreated cases into strata that are similar with respect to the distribution of covariates based on the propensity score distribution (Guo & Fraser, 2015). Because the estimated propensity score is a continuous variable, strata can be defined by the range of propensity scores – e.g., quintiles (5 strata), quartiles (4 strata) or deciles (10 strata). The larger the number of strata, the lower the variance (and potentially higher bias). Rosenbaum and Rubin (1984) recommended using quintiles as this number of strata was shown to remove about 90% of bias (Cochran, 1968). In reality, the number of strata is limited by the area of common support because each stratum must have at least one treated and one untreated case. Common support within each stratum may be improved with a reduced number of strata; however, bias removal may also be reduced if the treatment and control groups within a stratum differ substantially in sample size (Leite, 2017).

Two approaches can be used to estimate treatment effects through propensity score stratification:

(1) pooled strata-specific treatment effects: the difference in the averaged outcome between the treated and untreated groups will first be computed for each quintile and then summed across the five quintiles to estimate the overall treatment effect (Rosenbaum & Rubin, 1984);

(2) marginal mean weighting through stratification (MMWS): strata-based weights are created to "adjust for the difference between the observed proportions of treated and untreated units within strata and the proportions that would be obtained if randomized treatment assignment was used" (Leite, 2017).

The major difference between pooled strata-specific treatment effects and MMWS lies in the requirement for covariate balance: the former requires covariate balance within each stratum while the latter requires only marginal covariate balance. In the case of small stratum sizes, within-stratum covariate balance is more difficult to obtain and may lead to unreliable within-stratum group means for each covariate (Leite, 2017). Therefore, MMWS is preferred over pooled strata-specific treatment effects in estimating treatment effects through stratification.

- Weighting

Propensity score weighting consists of using weights that are derived directly from propensity scores to adjust the distribution of confounding variables so that they are similar for treated and untreated groups (Leite, 2017). Similar to sampling weights that adjust for bias due to over-sampling of cases with certain characteristics, propensity score weights adjust for over-selection of cases with certain characteristics to the treatment and control groups. Propensity score weighting is also known as inverse probability-of-treatment weighting (IPTW) for estimating the ATE -- the weights for treated cases are the inverse of the propensity score, i.e., $1/PS_i$; while the weights for untreated cases are the inverse of one minus the propensity score, i.e., $1/(1- PS_i)$. IPTW aims to adjust the sample so it is representative of a "pseudo-population" in which pre-treatment variables are balanced (Robins, Hernan & Brumback, 2000). In the context of multiple treatments, weights can be defined as the inverse of the generalized propensity score, or the inverse of the probabilities of a treatment condition for individuals who received that treatment.

The use of propensity score-based weights to reduce bias is not unique to propensity score weighting. For instance, variable ratio greedy matching without replacement can be viewed as a strategy to define weights where treated cases receive a weight of 1, matched untreated cases receive a weight equal to the inverse of the number of matches, an unmatched cases receive a weight of 0 (Leite, 2017). A key difference between these two methods lies in the preservation of sample size: most forms of matching result in discarding some unmatched cases from the analysis while propensity score weighting uses all the cases and thus helps to maintain statistical power in detecting treatment effects.

Note that unlike matching and stratification, propensity score weighting does not have a precise common support requirement and may still be able to obtain adequate covariate balance even if the distributions of propensity scores show little overlap (Ridgeway et al., 2017). However, poor common support may result in zero or extreme weights that inflate standard errors of treatment effects estimates and decrease power (Leite, 2017).

Last but not least, doubly robust estimation of treatment effects can be obtained with propensity score weighting. The doubly robust estimator consists of regression estimation of propensity score weighted means using the propensity score as a covariate and a comparison of the weighted means of the outcome between treated and untreated groups (Schafer & Kang, 2008). In the context of multiple treatments, propensity score weighting can be obtained with regression modelling of the relationship between covariates and the outcome for each treatment (Robins, Hernan & Brumback, 2000).

The combination of outcome regression (regression model) with weighting by propensity scores (propensity score model) is robust to misspecification of one (but not both) of these models (Bang & Robin, 2005; Tsiatis & Davidian, 2007). Lunceford and Davidian (2004) demonstrated that the doubly robust estimator performs better than stratification and IPTW. However, doubly robust estimator can still be biased if both the regression model and the propensity score model are misspecified (Funk et al., 2011).

2.4 Strengths and Limitations of Propensity Score Methods

Propensity score methods are particularly desirable for the analysis of large-scale survey data where a large number of confounding background characteristics need to be taken into consideration. As the number of covariates increases, the number of possible covariate values to be matched increases exponentially. With just 20 binary covariates, there are 2^{20} or about a million covariate patterns – it is very difficult to find an appropriate match for each treated participant with respect to all covariates. Propensity score methods solve this problem "by reducing the entire collection of background characteristics to a single composite characteristic" that appropriately summarizes the relationship between covariates and the treatment assignment (Austin, 2011). The reduction from many observed covariates to a single composite score not only greatly simplifies the analysis of large observational data but also allows a straightforward estimation of the average treatment effect.

Despite the broad utility of propensity score methods to estimate the ATE, it is important to keep in mind one critical distinction: propensity score methods rely on the principle that bias can be reduced by balancing the distribution of observed covariates (and the components of unobserved covariates that are correlated with the observed covariates) between the control and treatment groups, but not the distribution of unobserved ones. Therefore, hidden bias may remain due to unobserved confounders that are not correlated with the observed covariates but are correlated with the outcome. This is an inherent limitation of observational studies compared with RCTs "where the randomization tends to balance the distribution of all covariates, observed and unobserved" (Rubin, 1997).

Chapter 3. Research Design & Methods

3.1 Sample of Full-time Professionals from the 14 Selected Countries

The target population of this study comprises full-time professionals in all participating countries who have taken the computer-based PS-TRE assessment. As mentioned in Section 1.2, four countries -- Cyprus, France, Italy and Spain -- did not participate in the PIAAC problem-solving assessment and are thus dropped from this analysis. What is more, the sample for the Russian Federation excluded "the population residing in the Moscow municipal area" (OECD, 2013c). To ensure that this sample is a good representation of the target population across countries, we do not include the Russian Federation in this study. Lastly, data for Australia are not available on the OECD Public Use Data File.

As for the operational definition of full-time professionals, two criteria should apply here: full-time employees and professionals. Although there is no international standard regarding the minimum number of hours required to be counted as full-time employed, companies across the globe commonly require at least 30 hours of service per week in the main job. However, in household surveys like PIAAC, respondents are usually asked to self-identify their work status using the hours cut-off considered most suitable for the country concerned (Hanushek et al., 2011). In other words, respondents' self-reported employment status is used to distinguish full-time employees from part-time and unemployed individuals. For this study, full-time employees refer to respondents who identified their current status as "full-time employed (self-employed or employee)" in the PIAAC background questionnaire. Canada is the only country that did not include working hours or employment status; therefore, we have to leave the Canadian sample out of the analysis.

Furthermore, the focal sample is restricted to those who are full-time employed (self-employed or employee) in one of the six occupational sub-categories comprising the over-arching category of "Professionals" as defined by the ISCO-08 framework. The ISCO-08 framework split occupations into 10 major groups according to the tasks and duties undertaken in the job. Each major group is further organized into sub-groups based on the skill level and specialization required to competently perform the tasks and duties of the occupations. Table 4 lists the 10 major groups and the six sub-categories falling under the major group of "Professionals". All but three -- Austria, Estonia and Finland -- of the countries remaining in the sample have collected information of respondents' job at the sub-group level.

Table 4. The ISCO-08 Framework & Its Occupational Classifications of "Professionals"

Major Group 0 "Armed Forces Occupations"	
Major Group 1 "Managers"	
Major Group 2 "Professionals"	21 = Science and Engineering Professionals
	22 = Health Professionals
	23 = Teaching Professionals
	24 = Business and Administration Professionals
	25 = Information and Communications Technology Professionals
	26 = Legal, Social and Cultural Professionals
Major Group 3 "Technicians & Associate Professionals"	
Major Group 4 "Clerical Support Workers"	
Major Group 5 "Service & Sales Workers"	
Major Group 6 "Skills Agricultural, Forestry & Fishery Workers"	
Major Group 7 "Craft & Related Trades Workers"	
Major Group 8 "Plant & Machine Operators & Assemblers"	
Major Group 9 "Elementary Occupations"	

As noted earlier, restricting the analytical sample to full-time professionals yields a relatively homogeneous pool of workers in terms of time commitment (full-time vs. part-time) and skill level requirements (professionals vs. non-professionals). More specifically, this strategy enables us to more effectively isolate the relationships between skills proficiency and skills use at work from the moderating effects of employment status and job requirements. In a competitive labor market, higher level skills are systematically related to higher probabilities of employment (Hanushek et al., 2011). What's more, adults with higher level skills are more likely to land jobs in high-skill occupations, which tend to provide more opportunities for the maintenance and development of these skills over their lifetime. The *First Results from the Survey of Adult Skills* (OECD, 2013c, Ch. 3) examined the differences in skills proficiency among adults who work in low- and high-skilled occupations and confirmed that differences in skills proficiency are clearly associated with occupational differences in skill requirements. Furthermore, it is found that "the type of jobs held by workers is the single most important factor determining how individuals use their skills at work" (OECD, 2013c, Ch. 4).

Last but not least, to improve the policy relevance of the estimates, we eliminated the youngest cohort (aged 16-24) from this analysis due to small sample sizes. Among the 14 countries remaining in the sample, merely 3.6% of full-time professionals were under the age of 25. Such a low percentage is understandable because most people in this age band have not completed their formal education and may not be eligible to work full time as professionals. As a result, the final sample contains 8,535 full-time professionals from the 14 selected countries who were at least 25 years old at the time of the survey. Table 5 demonstrates the changes in sample size by country as more filters were added. As it turns out, on average across the 14 selected countries, full-time professionals aged 25 and above comprised just 9.9% of the PIAAC sample (before applying sampling weights)– the percentages are highest amongst Scandinavian countries -- Denmark, Sweden and Norway -- and lowest in the so-called Visegrád Group -- Poland, the Czech Republic and Slovak Republic.

Table 5. Summary of Sample Size Procedures by Country (absolute number & percentage)

Country	Overall	Full-time only	Full-time Pro.	%	Full-time Pro. Aged 25+	%	Full-time Asso. Aged 25+	%
Belgium	5,463	2,494	551	*10.1%*	521	*9.5%*	366	*6.7%*
Czech	6,102	2,896	454	*7.4%*	436	*7.1%*	493	*8.1%*
Denmark	7,328	3,809	1,254	*17.1%*	1,246	*17%*	549	*7.5%*
Germany	5,465	2,563	579	*10.6%*	528	*9.7%*	400	*7.3%*
Ireland	5,983	2,544	709	*11.9%*	688	*11.5%*	297	*5.0%*
Japan	5,278	2,743	468	*8.9%*	447	*8.5%*	474	*9.0%*
South Korea	6,667	3,575	531	*8.0%*	510	*7.6%*	425	*6.4%*
The Netherlands	5,170	2,078	494	*9.6%*	473	*9.1%*	339	*6.6%*
Norway	5,128	2,897	687	*13.4%*	676	*13.2%*	515	*10.0%*
Poland	9,366	3,919	604	*6.4%*	518	*5.5%*	367	*3.9%*
Slovak	5,723	2,953	450	*7.9%*	436	*7.6%*	486	*8.5%*
Sweden	4,469	2,488	716	*16.0%*	710	*15.9%*	488	*10.9%*
United Kingdom	8,892	4,086	801	*9.0%*	766	*8.6%*	509	*5.7%*
United States	5,010	2,532	606	*12.1%*	580	*11.6%*	438	*8.7%*
Total	**86,044**	**41,577**	**8,851**	***10.3%***	**8,535**	***9.9%***	**6,146**	***7.1%***

3.2 Selection of Covariates

A nested sequence of logistic regression models will be introduced to investigate the increase (or lack thereof) in explained variance of problem solving proficiency as more predictors are included. In view of the shortage of well-established assessments for adults' problem solving skills before PIAAC, two sets of parallel studies will be conducted to complement the main study on problem solving: the analysis of two foundational cognitive skills (i.e., Literacy and Numeracy). Comparing results from the main study with those from the parallel studies will help us better understand similarities and differences between foundational cognitive skills and problem solving in technology-rich environments (PS-TRE).

Previous research using large-scale assessment data across countries has shown that formal education plays a key role in developing foundation skills (OECD & Statistics Canada, 2000, 2005; OECD, 2013c). However, a growing body of evidence suggests that educational attainment does not fully explain skill differences amongst adults, and other factors must also play an important role (Massing & Schneider, 2017). For example, in ALL and PIAAC, there are substantial within- and cross-country variations in average proficiency levels at the same broad level of educational attainment (as defined by the International Standard Classification of Education 2011) (OECD & Statistics Canada, 2005; OECD, 2013c). The *First Results from the Survey of Adult Skills* (OECD, 2013c, Chapters 3 & 5) looked closely into a number of possible factors related to skills proficiency when controlling for educational qualifications. It was found that socio-economic background (proxied by parents' education) "has a strong impact on skills in some countries" and the relationship "is much weaker among younger adults than among older adults". More recently, in an attempt to "quantify the length of the shadow cast by family background", Braun (2018) found that "the well-known relationship between parental education and cognitive skills persists even when controlling for educational attainment".

A thematic review on potential explanatory variables of full-time professionals' PS-TRE proficiency yielded five blocks of factors that will be entered into the logistic regression models in sequence. The rationale behind selecting these variables is elaborated in Section 2.1. As indicated in Table 6, these selected variables are organized by socio-demographic background, occupational category, use of key information-processing skills at work (and at home), use of generic workplace skills/job-related activities, as well as AET participation. Proficiency scores in problem solving in technology-rich environments (PS-TRE) is the outcome of interest for the main study; while Literacy and Numeracy proficiency levels are the outcomes for the two sets of parallel studies respectively. Last but not least, to mitigate the effects of selection and confounding bias that is endemic to observational studies, propensity score methods will be employed to better capture the factors that lead to the non-randomized allocation of AET participation among full-time professionals.

Table 6. Overview of Predictor Variables Entered into the Logistic Regression Models

Clusters	***Explanatory Variables***
1) Socio-Demographic Background	*Age (in 10-year bands), Gender, Family Background (proxied by Parental Education), Educational Attainment*
2) Occupational Categories	*Science & Engineering, Health, Teaching, Business & Administration, Information & Communications Technology (ICT) Legal, Social & Cultural*
3.1) Use of Key Information-Processing Skills at Work	*Reading, Writing, Numeracy, ICT*[18]
3.2) Use of Key Information-Processing Skills at Home	*Reading, Writing, Numeracy, ICT*
4) Use of Generic Workplace Skills / Job-Related Activities	*Influencing, Planning, Task Discretion, Learning at work*
5) AET Participation	*Formal AET Programs Non-formal AET Programs*

Outcome Variable:
Proficiency Scores in Problem Solving in Technology-Rich Environments (PS-TRE)
(Literacy & Numeracy Proficiency Levels for parallel studies)

Considering the complexity of adult working life, skills used at work are sorted into two groups: key information processing skills and generic workplace skills/job-related activities. For each skill, the intensity of usage is measured indirectly through a set of self-reported questions on a 0-4 scale. Possible answers to these questions are ordered "whereby consecutive alternatives always indicate a higher frequency of performing a certain task: 0 = never used; 1 = less than

[18] ICT stands for Information and Communication Technologies. In addition to analytic problem solving skills per se, the PIAAC PS-TRE domain emphasizes on "computer literacy" skills (i.e. the capacity to use ICT tools and applications) required to solve "information problems". Hence the mastery of foundation ICT skills is a prerequisite for PS-TRE proficiency (OECD, 2012a, p. 39).

once a month; 2 = less than once a week but at least once a month; 3 = at least once a week but not every day and 4 = every day.

Because all of the aforementioned skills can be linked to more than one item in the background questionnaire, PIAAC used item response theory (IRT) (OECD 2013a), to derive a continuous latent scale for each type of skill used at work (and at home). IRT modeling, more specifically the generalized partial credit model (GPCM), is an appropriate statistical method to derive indicators of skills use from groups of multiple choice questions "where the item responses are contained in two or more ordered categories". Items associated with a given skill used at work are grouped together and the intensity of usage is estimated. The resulting continuous, one-dimensional scale accounts for a substantial proportion of the covariances among the item responses. Respondents with a higher level on the derived scale have "a higher probability of frequently performing the task detailed in a given item" (OECDb, p. 44).

As per OECD documentation (OECD, 2013b, 2013c), a total of twelve skills use indicators were created (see Appendix C). The application of these criteria in subscale retention determinations is formally described in Chapter 20.5.4-20.5.6 of the PIAAC technical report (OECD, 2013a). After a thorough evaluation of their psychometric qualities, only eight were retained for this analysis (see Table 2). The first four indicators refer to key information-processing skills -- reading, writing, numeracy and ICT skills[19]; the remaining four indicators correspond to general skills -- influencing, planning, task discretion and learning at work[20].

---

[19] The PIAAC background questionnaire only has two items for the problem solving subscale. It was flagged as problematic due to large between-country differences (≥ the criterion .25 logits) (OECD, 2013, Ch. 20, p. 17). To address concerns about reliability (Eisinga, Te Grotenhuis, & Pelzer, 2013), this two-item scale will not be included in this analysis.

[20] Among the seven indicators of workplace skills, co-operation and physical skills subscales were flagged as problematic both for low reliability and between country differences. Self-organization subscale was strongly

To make meaningful comparisons between different types of skills used at work, all indicators have been standardized to have mean equal to 2 and standard deviation equal to 1 across the entire survey sample in all 23 participating countries (appropriately weighted) (OECD, 2013b, 2013c). As a result, at least 90% of surveyed respondents had indicator scores between 0 and 4, "whereby values approaching 0 suggest a low frequency of use and values approaching 4 suggest a high frequency" (OECD, 2013b, p. 44). For each type of skill, the average level and the inter-quartile range of usage at work will be computed and compared across countries and occupational categories. Within each country and occupational category, the average usage level of each skill will be further examined by demographic subgroups (i.e. gender and age). Skills use indicators demonstrating significant between-group differences will be identified and potential explanations will be discussed at the national and occupational level.

correlated with the three-item scale of planning, indicating redundant information. These scales were excluded from this analysis.

3.3 Research Questions

RQ1a. What are the distributions of full-time professionals' PS-TRE proficiency scores across the 14 selected countries that participated in the first round of PIAAC?

RQ1b. Within each country, how do the distributions of full-time professionals' PS-TRE proficiency scores vary by gender and age?

RQ1c. How do the relationships between full-time professionals' PS-TRE proficiency scores and their proficiency scores in Literacy and Numeracy vary across the 14 selected countries?

RQ1d. How do the distributions of full-time professionals' PS-TRE proficiency scores compare to those of full-time associates in the aforementioned occupational sub-categories (except Teaching)?

The answer to RQ1a describes full-time professionals' levels of proficiency in PS-TRE and compares distributions of proficiency levels across and within countries. For each of the 14 selected countries, the mean (and standard deviation) of full-time professionals' PS-TRE scores will be calculated and compared with those of full-time workers in other occupational categories, i.e., full-time associate professionals.

In accordance with the theoretical framework for factors related to PS-TRE, we hypothesize that full-time professionals, as a result of their educational attainment and job requirements, should demonstrate higher levels of proficiency in PS-TRE than full-time workers in other occupational categories. Weighted t-tests will be performed to compare the average PS-TRE scores of full-time professionals with those of full-time associate professionals, both within and across countries. Box and whisker plots will be drawn to visualize full-time professionals' comparative advantage (if there is any) in solving problems in technology-rich environments.

Countries showing significant between–group differences (i.e., full-time professionals vs. full-time associate professionals) in the weighted means of PS-TRE scores will be ranked based on the effect size (i.e, Cohen's *d*), rather than the raw mean difference. Defined as "a quantitative measure of the strength of a phenomenon", the effect size corresponds to a standardized mean difference which is actually the difference between two means split by the pooled standard deviation for two independent samples (weighted by the respective sample sizes) (Kelley & Preacher, 2012).

Similarly, we will compute the means (and standard deviations) of full-time professionals' Literacy and Numeracy scores and compare them with those of full-time associate professionals. For each country, weighted t-tests and box plots will again be used to identify full-time professionals' comparative advantage (or lack thereof) in traditional cognitive skills. As stated in the *First Results from the Survey of Adult Skills* (OECD, 2013c), "high levels of proficiency in literacy and numeracy go hand in hand with high levels of proficiency in problem solving in digital environments" (p. 96), we hypothesize that full-time professionals will demonstrate an advantage over full-time workers in other occupational categories for all three key information-processing skills.

Based on the statement that "high levels of proficiency in literacy and numeracy go hand in hand with high levels of proficiency in problem solving in digital environments" (OECD, 2013c), the third subset of research question one involves comparing full-time professionals' distributions of PS-TRE proficiency to their distributions of proficiencies in Literacy and Numeracy, if the latter are reported on an approximately comparable scale. To facilitate comparisons across skill domains, Holzer and Lerman (2015) re-organize the original reporting scales (in proficiency scores) into a simplified three-level scale that applies to PS-TRE as well as

Literacy and Numeracy: Low, Medium and High Proficiency. Please refer to p. 13-14 regarding how this simplified scale defines low, medium and high proficiency differently for PS-TRE, Literacy and Numeracy. The distributions of full-time professionals' proficiency in each skill domain -- in terms of percentages scoring at low, medium and high proficiency levels -- will be obtained and compared across countries. To assist comprehensive comparisons between PS-TRE and traditional cognitive skills, clustered bar charts will be utilized. Clustered bar charts allow for the display of two categorical variables, in this case, skill domains and proficiency levels. For each country, percentages across the three proficiency levels will be clustered by skill domain.

However, the massive information produced by these charts will make it very difficult to describe cross- and within- country patterns in a straightforward way. Guided by the hypothesis that full-time professionals should represent the most competent group of the adult population, our discussions will focus on the proportion of "strong" performers in each skill domain, i.e., those who have achieved high proficiency in Literacy, Numeracy and PS-TRE respectively. Based on the number of domains in which they are "highly proficient", we can classify out sample into four groups with 0 indicating failure to achieve high proficiency in any key information-processing skills and 3 indicating success in achieving high proficiency in all three skill domains. We can then rank the 14 selected countries using different criteria; for example, the percentage of "strong" performers in PS-TRE alone, in both Literacy and Numeracy, and in PS-TRE, Literacy and Numeracy altogether. If the "hand in hand" relationship (OECD, 2013c, p. 96) among key information-processing skills remains true for full-time professionals, we should come up with more or less the same rankings regardless of which criterion is followed.

Within each of the 14 selected countries that participated in PIAAC, we will compare the distribution of "strong" performers (in proficiency scores) in each skill domain by demographic subgroups. Cross tabulations will be deployed to present the multivariate distribution of "strong" performers by gender-and-age subgroups in Literacy, Numeracy and PS-TRE respectively. The hypothesis is that other things being equal, certain demographic subgroups (e.g., males or millennials) tend to have higher proficiency in managing information in the digital era than others (e.g., females or baby boomers).

To determine if there is an interaction between gender and age cohort on the scale scores of key information-processing skills, a two-way analysis of variance (ANOVA) will be conducted for each country. If a statistically significant interaction is present, the association between gender and the scale score varies by age cohort. Additional analysis procedures are needed to look at the association (between gender and the scale score) separately for each age cohort (i.e., simple main effects). In our example, this would involve examining the mean difference in the PS-TRE scale score between genders for each age cohort. Note that there are 4 simple main effects for a 2x4 factorial design – 1 for the gender factor and 3 for the age factor. If the interaction term is not significant but main effects are, Tukey's HSD test can be relied on to identify where the significant differences lie between/among gender or age groups. Tukey's HSD (honest significant difference) test is a single-step multiple comparison procedure and statistical test that can be used on raw data or in conjunction with an ANOVA post-hoc analysis to find group means significantly different from each other (Tukey, 1949). To graphically illustrate the findings, the average score for each combination of gender and age cohort will be plotted in a multi-series line chart (each line represents a gender) per skill domain and per participating country.

Besides comparing mean differences in the scale scores of key information-processing skills among groups defined by gender and age cohort within each country, it is worth paying attention to overall gender and age gaps across countries. In this study, the age gap is operationally defined as the difference between the youngest and the old age cohorts on each scale (Literacy, Numeracy or PS-TRE). Three graphs will be constructed to facilitate cross-country comparisons for each skill domain. The x-axis will designate gender gaps and y-axis age gaps so that each country is represented by a point in the plane. Outlying data points (i.e., countries) will be identified and possible explanations will be proposed.

Moreover, we can compare the reversed country rankings of gender or age gap (set as x-axis) with that based on the scale score of key information-processing skills in each skill domain (set as y-axis). Country rankings based on the width of gender or age gap are reversed so that a higher ranking number indicates a narrower gender or age gap and thus more equity. This way, the position of a country in the plane indicates the "compatibility" between gender/age equity and skills proficiency (i.e., the equity-proficiency map). Visually speaking, quadrant 1 features (model) countries that have achieved high proficiency and high gender/age equity. By contrast, quadrant 3 includes countries with lower rankings in both skills proficiency and gender/age equity. Quadrants 2 and 4 represent countries where a proficiency-equity trade-off is observed.

Apart from cross-country variation, occupational category is another important factor that may account for proficiency gaps among gender-and-age subgroups. Depending on the nature of the job, some professions may have a higher entry requirement for certain skills. For instance, one must demonstrate adequate mastery of computer techniques in order to get into the IT industry; while reading and writing proficiencies are much sought after in the recruitment of legal, social and cultural professionals. To minimize the moderating effects of occupational selection (due to differences in skill requirements), it is necessary to describe full-time professionals' patterns of skills proficiency by occupational category. Within each of the six occupational sub-categories comprising "professionals", we will observe and compare the distributions of "strong" performers (in proficiency scores) in each skill domain by gender-and-age subgroups within each country. A three-way ANOVA with occupational category as the third factor will be conducted where a significant three-way interaction means that one, or more, two-way interactions differ across the levels of a third factor. For example, we can examine how the gender*age interaction changes across occupational categories for each country. Note that to reduce the number of cells (and thus increase the sample size in each cell), the four age groups may be combined into two: age 25-44 and age 45-65.

Notwithstanding the pervasive skill gaps found in the literature, we suspect that even within the same country, some professions are more likely to show wider gender or age gaps in skills proficiency than others. To capture the "age-related cognitive maturation and decline" [21] (OECD, 2013c, p. 105) for each gender, the average PS-TRE score for each age cohort will be plotted in a double-series line chart (each line represents a gender) per occupational category and per country. For each country, the six occupational categories will be (reversely) ranked on their gender or age gap (set as x-axis) and presented in a coordinate plane with rankings based on the PS-TRE scale score (set as y-axis). Occupations in quadrant 1 of the "equity-proficiency map" (i.e., high equity & high proficiency) will be identified and compared to those in quadrant 3 (i.e., low equity & low proficiency). Occupational categories showing a trade-off between gender/age equity and skills proficiency will also be examined with acknowledgement of the inherent structural differences between professional fields.

[21] Rather than a developmental curve of skills proficiency of an age cohort at different points in time, the PIAAC data offers a snapshot of skills proficiency of adults of different age cohorts at the time of the survey (OECD, 2013c, p.105).

Compared to cross-country analyses, interpretations of gender- or age-related proficiency gaps at the occupational level may be subject to more constraints. Other than the moderating effects of occupational selection, the particular roles and responsibilities aligned with specific job positions may also account for variations in full-time professionals' proficiency in key information-processing skills and readiness to learn. In research question two, we will adopt a use-oriented approach to investigate variations in skills proficiency related to socio-demographic background factors, occupational categories and more importantly, different types and intensitites of skills use at work.

RQ2. For full-time professionals within each country, how do their socio-demographic background factors, occupational categories, and the types and intensities of skills use at work (and at home), relate to their probabilities of participating in the different formats of AET programs?

Thanks to the JRA module, the PIAAC background questionnaire collected a large amount of information about the frequency with which a wide range of skills are used in order to perform specific work tasks. As mentioned earlier, eight IRT-derived skill use indicators were retained for this analysis. Each has "been standardized to have mean equal to 2 and standard deviation equal to 1 across the pooled sample of respondents in all countries (appropriately weighted)" (OECD, 2013b, p. 44). For full-time professionals within each country, we will compute and compare the average level of use across the eight skills. A line chart will make it easy to identify the most intensely used skill(s) at the national level. What is more, cluster analysis of the PIAAC data suggested that key information-processing skills tend to be used together at work, while workplace skills are not (with the exception of influencing skills which tend to be associated with reading, writing and problem-solving skills) (OECD, 2013c, p.147). Summarizing the clustering (if there is any) and the extent of skills use accordingly will offer a more fine-grained picture of the differences in skill demands both within and across countries.

As a side note, the use of generic workplace skills may be strongly associated with the conditions and demands that full-time professionals encounter in their jobs (thus the name job-related activities). Correspondingly, different occupational sub-categories of “Professionals” should demonstrate different patterns for the use of these skills in the workplace. Also considered as indicators of High-Performance Work Practices (HPWP)[22] for work settings, the intensity of using generic workplace skills indicates how widespread HPWP are across countries and occupational categories. Additional analyses conducted for the OECD skills studies confirmed that workplace skills are “closely and positively related to” the use of information-processing skills at work (OECD, 2016, p. 28). In this regard, we should not be surprised if the distributions of workplace skills by socio-demographic background and occupational category resemble those of key information-processing skills.

Socio-demographic background should also play a role in explaining full-time professionals’ variations in skills use at work. According to the *First Results from the Survey of Adult Skills* (OECD, 2013c), men and prime-age workers (i.e., aged 25-54) use key information-processing skills at work more frequently than women and those aged 55-65 in most participating countries (p. 149 & p. 152). However, it has been found that differences in skills use by gender decreased when skills proficiency and occupational categories are taken into

[22] HPWP includes two aspects: work organization and management practices. The Survey of Adult Skills collects information on a number of job-related activities that are often associated with the work organization aspect of HPWP -- how often they instruct, teach or train other people (i.e., influencing); how often they organize their own time and plan their own activities (i.e., planning); whether workers have any flexibility in deciding on the sequence of tasks they perform (i.e., task discretion), and whether they participated in education/training in the previous 12 months (i.e., learning at work) (OECD, 2016, p.113).

account (p. 149). A similar pattern applies to age cohorts as well (p. 152). Variations in skills use among gender and age groups may be a joint consequence of the national employment policy and individual career choice. For instance, women are more likely to sort themselves into jobs that presumably "require less investment in human capital during the period of childrearing" (OECD, 2013c, p. 182) and millennials are more attracted to the so-called sunrise industries in information and communications technology than baby boomers. Therefore, it is necessary to look at the demographic distribution of skills use by country and occupational category. In most (if not all) participating countries and professional fields, we expect to see an interaction between gender and age cohort in the use of skills at work. More specifically, gender differences in skills use at work are expected be smaller among younger workers as more and more millennial women are entering previously male-dominated fields (e.g., STEM industries) and undertaking the same roles and responsibilities as their male counterparts in the workplace.

Next, we will focus on modeling the relationships between predictors (i.e., socio-demographic background and skills use at work) and the outcome (i.e., full-time professionals' probabilities of participating in the different formats of AET programs). Since the outcome variable of research question two consists of three formats of AET participation without any kind of natural order (i.e., nominal variable with three categories), multinomial logistic regression is used to predict the probability of each categorical membership -- Formal and Non-formal AET participation (vs. None). As the name suggests, multinomial logistic regression is an extension of binomial logistic regression that (1) allows for more than two categories of the outcome variable and (2) models the log odds of an outcome category (vs. the reference) as a linear combination of the predictor variables. As with other types of regression, multinomial logistic regression can have categorical and/or continuous predictor variables.

Advantages of Multinomial Logistic Regression

Before delving into the details of building multinomial logistic regression models, it is necessary to acknowledge that there are several analysis methods that can also be applied to research question two. However, a comprehensive comparison of these methods suggest that multinomial logistic regression is the most appropriate analytic approach for the research question at hand:

(1) We can choose to run binomial logistic regression multiple times, one for each pair of outcome categories. In this case, there would be three pairs and thus three sets of analysis. The major problem with this method is that each analysis will be run on a different sample. Not to mention that the sum of the probabilities for the three formats of AET participation will likely be greater than 1.

(2) Alternatively, we can collapse the three outcome categories into two and then run a binomial logistic regression. However, collapsing categories may result in loss of information and even change the original research question to a different one.

(3) Multinomial probit regression which is similar to multinomial logistic regression with independent normal error terms. A noticeable difference between the two models is typically only seen in small samples because the probit function assumes a normal distribution of the probability of the event, whereas the logit function assumes a log distribution. Also note that for multinomial logistic regression, the sample size guidelines indicate a minimum of 10 cases per predictor variable (Schwab, 2002). In our model with 17 predictors, that is a minimum of 170 full-time professionals for each country. A direct reference back to Table 5 confirms that the minimum sample size requirement is met in all 14 countries.

(4) Multiple-group discriminant analysis which is a multivariate method for multinomial outcome variables. Discriminant analysis requires the assumptions of normality (i.e. normally distributed predictors), linearity, or homoscedasticity, whereas multinomial logistic regression does not have such assumptions.

Explanations of Multinomial Logistic Regression Models

To build logistic regression models, a nested sequence of logistic regression models will be introduced that includes blocks of control variables and indicators of skills use at work. To facilitate interpretation of the regression results, we adopt a block-wise entry method. Predictors are selected and grouped into blocks based on the theoretical framework introduced in Chapter 2. In general, predictors clustered in the same block will be inter-correlated. First, full-time professionals' socio-demographic background (i.e., gender, age, parental education and educational attainment) will be entered as baseline controls (Model 1). The purpose is to estimate each socio-demographic background factor's partial correlation with full-time professionals' probabilities of participating in the different formats of AET programs. Therefore, the baseline logistic regression equation (Model 1) will have the form:

$$\hat{Y} = a + b_1(\text{Male}) + b_2(\text{Age1}) + b_3(\text{Age2}) + b_4(\text{Age3}) + b_5(\text{ParEd1}) + b_6(\text{ParEd2}) + b_7(\text{EduAtt1}) + b_8(\text{EduAtt2})$$

The outcome is a three-level categorical variable indicating the three different formats of AET participation (reference category: None). This baseline model will output two rows of coefficients (and standard errors) – the first row comparing Non-formal AET participation to None and the second row comparing Formal AET participation to None. Note that in this specific example, all the independent variables are categorical and will be encoded using binary coding which compares each level of a variable to the omitted (reference) level. That is, for all but one of the levels of the categorical variable, a binary indicator is create that has a value of one for observations at that level and zero for all others. In the "EduAtt" variable example, the first indicator "EduAtt1" has a value of one for individuals with less than a Bachelor's degree, and zero for all other observations. Likewise, we create "EduAtt2" to be one for individuals with a Bachelor's degree and zero otherwise. The level of the categorical variable that is coded as zero in all the binary indicators is the reference level, or the level to which all of the other levels are compared. For the "EduAtt" variable, we select Master/Research degree as the reference level. Table 7 displays how the binary indicators are defined for each socio-demographic background factors in the baseline model (Model 1):

Table 7. Binary Coding for Socio-Demographic Background Factors (Model 1)

Original Variable	Reference Level	Binary Indicator Created
Gender	*Male*	Female (1 = Female; 0 = Male)
Age	*Age 55-65*	Age1 (1 = Age 25-34; 0 = Otherwise) Age2 (1 = Age 35-44; 0 = Otherwise) Age3 (1 = Age 45-54; 0 = Otherwise)
Parental Education (ParEd)	*At least one parent has attained tertiary*	ParEd1 (1 = Neither parent has attained upper secondary; 0 = Otherwise) ParEd2 (1 = At least one parent has attained secondary & post-secondary, non-tertiary; 0 = Otherwise)
Educational Attainment (EduAtt)	*Tertiary – Master/Research Degree*	EduAtt1 (1 = Less than Tertiary – Bachelor Degree; 0 = Otherwise EduAtt2 (1= Tertiary – Bachelor Degree; 0 = Otherwise)

Model 2 relates to the moderating effects of occupational choice on the associations between socio-demographic background factors and full-time professionals' probabilities of participating in the different formats of AET programs. In order to control for the ubiquitous moderating effects of occupational choice, it is necessary to take into consideration full-time professionals' occupational information. Please refer to Table 4 for a complete list of the six occupational sub-categories as defined by the ISCO-08 framework. Correspondingly, five dummy variables (reference category: Legal, Social & Cultural) will be added to Model 2 with 1 indicating working in that specific occupational sub-category whereas 0 indicating otherwise.

Model 3 and Model 4 will each include four IRT-derived continuous indicators of skills use in full-time professionals' job context, for key information-processing and generic workplace skills/job-related activities respectively. See Table 2 for a complete list of the eight indicators of skills use at work, as well as their corresponding sample tasks**.** Note that in constructing nested regression models, the order of entry has an impact on which variables will be selected: those that are entered in the earlier stages have a better chance of being retained than those entered at later stages. For this reason, we purposefully adopt such an order so that the distinct explanatory role of workplace skills/job-related activities (if there is any) will not be disguised by that of key information-processing skills. In other words, we want to distinguish the unique association between each type of generic workplace skills/job-related activities and full-time professionals' probabilities of participating in the different formats of AET programs, from the confounding effects of socio-demographic background, occupational category as well as the use of key information-processing skills at work. PIAAC's tasked-based JRA module also facilitates our exploratory effort to match the use of generic workplace skills/job-related activities (as an indicator of labor demand) with full-time professionals' AET participation and in research question three, full-time professionals' PS-TRE proficiency scores (as a key measure of labor supply).

Interpretations of Multinomial Logistic Regression Coefficients

At the center of multinomial logistic regression analysis is the task of estimating the logistic regression coefficient for each predictor, and for each alternative category of the outcome. By "alternative", we mean all the non-reference outcome categories. The logistic regression coefficient for a given predictor, or B, is the expected amount of change in the logit associated with one unit change in this specific predictor, after partialling out the variability due to the other variables in the model. The logit is what is being predicted which is the log odds of being in one alternative outcome category vs. the reference. The closer a logistic regression coefficient is to 0, the less influence the corresponding predictor has in predicting the log odds of being in the specified alternative category vs. the reference. That is to say, for a logistic regression coefficient, if the 99% confidence interval includes 0, the logistic regression coefficients are not significantly different from 0 and the corresponding predictor is not a significant predictor. Note that when a set of statistical inferences are considered simultaneously, some will have p values less than .05 purely by chance, even if all the null hypotheses are true. To account for the problem of multiple testing or multiplicity, we switch to a stricter significance threshold for individual testing, so as to compensate for the number of inferences being made. If not otherwise specified, the rest of the study will use a lower significance level .01, rather than .05, to determine whether to reject null hypothesis.

Because it is not intuitive to interpret these coefficients in terms of logits, the exponentiated values of the coefficients – i.e., odds ratios – are presented to facilitate simple interpretations. The exponentiated logistic regression coefficient for a given predictor, or Exp(B), is defined as a ratio of the odds for being in one alternative outcome category vs. the reference associated with one unit change in this specific predictor, after partialling out the variability due to the other variables in the model. Predictors with Exp(B) greater than 1 are expected to increase the odds of participating in the specified alternative category vs. the reference; whereas predictors with Exp(B) less than 1 are expected to do the opposite. Predictors with Exp(B) =1 are considered non-significant. Note that when considering logistic regression coefficients, we want the values to be significantly different from 0. If we exponentiate 0, we get Exp(0) = 1. Hence, this is two ways of saying the same thing.

To better explain the concepts of log odds and odds ratio, we will use the logistic regression coefficients and the exponentiated logistic regression coefficients for Educational Attainment (EduAtt) and ICTWORK from the Japanese results as an example. As stated in Table 7, we selected Master/Research degree as the reference level for "EduAtt" and created two binary indicators – "EduAtt1" = 1 for individuals with less than a Bachelor's degree and 0 otherwise and "EduAtt2" = 1 for individuals with a Bachelor's degree and 0 otherwise. While Learning at Work is a continuous variable standardized to have mean equal to 2 and standard deviation equal to 1. An important note is that we are using SPSS Statistics to run multinomial logistic regression and the default behavior in this software is for the last category (numerically) to be selected as the reference category. In the example below, this is full-time professionals who have obtained a Master/Research degree or above.

Table 8. Example from the Japanese Results of Multinomial Logistic Regression

	Coefficient of EduAtt1 = less than BA	Exp(B)	Coefficient of EduAtt2 = B.A.	Exp(B)	Coefficient of Learn@Work	Exp(B)
Non-formal vs. None	-1.77**	.17	-1.35**	.26	.67***	1.94
Formal vs. None	-2.76**	.06	-2.47**	.09	.80	2.23

** Significant at the .05 level, ** Significant at the .01 level, ***Significant at the .001 level*

As can be seen in Table 8, each row of coefficient values corresponds to a logit equation -- the first row compares Non-formal AET Participation to None (the reference) and the second row compares Formal AET Participation to None (the reference). If we consider the coefficients from the first row to be b_1 and the coefficients from the second row to be b_2, we can write two simple logit equations as follows. Note that other predictors are omitted in this example for simplicity.

$$ln(\frac{P(AET=Nonformal)}{P(AET=None)}) = b_{10} + b_{11}(EduAtt = 1) + b_{12}(EduAtt = 2) + b_{13}(ICTWORK) + \ldots$$

$$ln(\frac{P(AET=Formal)}{P(AET=None)}) = b_{20} + b_{21}(EduAtt = 1) + b_{22}(EduAtt = 2) + b_{23}(ICTWORK) + \ldots$$

Logistic Regression Coefficients of Educational Attainment:

Applying the first equation to the Japanese example, each estimated coefficient is the expected change in the log odds of participating in Non-formal AET programs (vs. None) for a unit change in the corresponding predictor variable, after partialling out the variability due to the other variables in the model. Each exponentiated coefficient is the ratio of two odds for a unit change in the corresponding predictor variable, after partialling out the variability due to the other variables in the model.

b_{11} = -1.77: After partialling out the variability due to the other variables in the model, the odds of participating in Non-formal AET programs for individuals with less than a Bachelor's degree (EduAtt1 = 1) over the odds of participating in Non-formal AET programs for individuals with a Master/Research degree (EduAtt1 = 0) is exp(-1.77) = .17. In terms of percent change, we can say that after partialling out the variability due to the other variables in the model, the odds of participating in Non-formal AET programs for those with less than a Bachelor's degree are 83% lower (i.e., .17 – 1 = -.83) than those with a Master's degree or above.

b_{12} = -1.35: After partialling out the variability due to the other variables in the model, the odds of participating in Non-formal AET programs for individuals with a Bachelor's degree (EduAtt2 = 1) over the odds of participating in Non-formal AET programs for individuals with a Master/Research degree (EduAtt2 = 0) is exp(-1.35) = .26. Or after partialling out the variability due to the other variables in the model, the odds of participating in Non-formal AET programs for those with a Bachelor's degree are 74% lower (i.e., .26 – 1 = -.74) than those with a Master's degree or above.

Likewise, the estimated coefficient of the second equation is the expect change in the log odds of participating in Formal AET programs (vs. None) for a unit change in the corresponding predictor variable, after partialling out the variability due to the other variables in the model.

b_{21} = -2.76: After partialling out the variability due to the other variables in the model., the odds of participating in Formal AET programs for individuals with less than a Bachelor's degree (EduAtt1 = 1) over the odds of participating in Formal AET programs for individuals with a Master/Research degree (EduAtt1 = 0) is exp(-2.76) = .06. In terms of percent change, after partialling out the variability due to the other variables in the model, the odds of participating in Formal AET programs for those with less than a Bachelor's degree are 94% lower (i.e., .06 – 1 = -.94) than those with a Master's degree or above.

b_{22} = -2.47: After partialling out the variability due to the other variables in the model, the odds of participating in Formal AET programs for individuals with a Bachelor's degree (EduAtt2 = 1) over the odds of participating in Formal AET programs for individuals with a Master/Research degree (EduAtt2 = 0) is exp(-2.47) = .09. Or after partialling out the variability due to the other variables in the model, the odds of participating in Formal AET programs for those with a Bachelor's degree are 91% lower (i.e., .09 – 1 = -.91) than those with a Master's degree or above.

Logistic Regression Coefficients of Learning at Work Skills:

b_{13} = .67: After partialling out the variability due to the other variables in the model, a one-unit increase in the predictor Learning at Work is associated with 95% increase in the odds of participating in Non-formal AET programs since exp(.67) = 1.95.

b_{23} = .80: After partialling out the variability due to the other variables in the model, a one-unit increase in the predictor Learning at Work is associated with 123% increase in the odds of participating in Formal AET programs since exp(.80) = 2.23. However, this coefficient is not significantly different from 0 ($p = .053$).

In sum, for two Japanese full-time professionals with the same predictor profile except for Education Attainment, a higher degree level is significantly associated with an increase in the odds of participating in both formats of AET programs, but more so for Formal AET participation. On the other hand, after partialling out the variability due to the other variables in the model, a higher intensity of using Learning at Work skills is significantly associated with an increase in the odds of participating in Non-formal AET programs but not Formal AET programs.

RQ3. Comparing full-time professionals estimated to have similar probabilities (i.e., propensity scores) of participation in a certain format of AET programs, do participants of certain format of AET programs score higher than non-participants as measured by PIAAC PS-TRE?

(Note: In RQ4, parallel analyses will be conducted with (1) full-time associate professionals' PS-TRE proficiency scores and (2) full-time professionals' Literacy and Numeracy scores as the outcome)

In PIAAC, AET participation is measured by a set of categorical variables, indicating participation in a variety of lifelong learning activities within the 12 months prior to the survey. To sort out information that captures the most popular trend of AET participation among full-time professionals, three dichotomous indicators are chosen for the analyses -- describing respondents' participation in None, Non-formal AET and Formal AET programs respectively. We will begin with an examination of the average participation rates for formal and non-formal AET programs separately across the 14 selected countries. On account of systematic national and institutional differences in education and training system, labor market and employment policy as well as the nature of professional practice, there are likely to be substantial cross-country variations in full-time professionals' choices of AET participation.

For each of the 14 selected countries, the goal of research question three is to shed light on (1) the relative popularity of the different formats of AET programs among the most privileged group of workers in the labor market; (2) individual (e.g. gender, age, parental education and educational attainment) and job characteristics (e.g., occupational categories, the types and intensities of skills use at work) accounting for variations in full-time professionals' participation in the different formats of AET programs and (3) the relationships between AET participation and PS-TRE proficiency scores after controlling for socio-demographic background factors, occupational categories, and the types and intensities of skills use at work. Although the hypothesis is that there should be a positive association between AET participation and proficiency scores measured by PIAAC (Sgobbi, 2014; Cegolon, 2016), the magnitudes and significance of the associations may vary widely across countries and domains. It is therefore necessary to first examine the distribution of AET participation rates across countries, especially for those where the association turned out to be negative or non-significant.

In order to obtain approximately unbiased estimates of the association between AET participation (either Formal or Non-formal) and full-time professionals' PS-TRE proficiency scores, we need to work on reducing selection bias due to the absence of random assignment in observational studies. As the first part of research question three suggests, there may exist some baseline differences between adults who chose to participate in (either Formal or Non-formal) AET programs and those who did not. These baseline differences in individual and job characteristics can be sources of pre-treatment heterogeneity between AET participants (either Formal or Non-formal) and non-participants. To "purge", in Murnane and Willett's term (2011), the estimated treatment effect of the impact of selection bias, each individual will have an estimated PS-TRE proficiency score as a result of participating in one of the three alternative AET programs – None, Non-formal and Formal -- accordingly denoted as PS-TRE$_{t1}$, PS-TRE$_{t2}$ and PS-TRE$_{t3}$. Note that for sake of easy comparison and consistent terminology, "did not participate in any AET program" is considered as a treatment condition here.

3.3.1 Estimation of Propensity Scores

It is worth noting that propensity scores can be defined in both RCTs and observational studies. In RCTs the true propensity score is known or can be constructed from the experimental design. In observational studies, however, the true propensity score is usually not known due to the absence of random assignment. Most often, the propensity score is estimated from a logistic regression model with treatment assignment (denoted as $Z = 1$) as the outcome and measured covariates (denoted as $X_1, X_2, X_3, \dots X_k$) as the predictors. Therefore, the propensity score is the predicted probability of treatment assignment given covariates (denoted as $Pr(Z = 1|X)$). According to Leite (2017), a basic logistic regression model for estimating propensity scores for individual i is as follows:

$$\text{logit}(Z_i = 1|X_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} \dots \beta_k X_{ki} \qquad \textbf{Equation 6}$$

where the logit is the log odds of the probability of receiving treatment Z:

$$logit(Z_i = 1|X_i) = log\left(\frac{Pr(Z_i = 1)}{1 - Pr(Z_i = 1)}\right) \qquad \textbf{Equation 7}$$

Because the strong ignorability of treatment assignment assumption requires that there be no omitted confounders, attempts should be made to include all true confounders, “which are covariates that affect the treatment assignment and the outcome” (Leite, 2017). Besides true confounders, outcome predictors that are unrelated to treatment assignment should also be included to increase the power to estimate the treatment effect.

Propensity scores are typically estimated to control for systematic differences between two groups (i.e., treatment vs. control) and it is not until recently that these techniques have been extended to situations where there are two or more treatment conditions (i.e., treatment A and treatment B). A classic example would be a physician may choose among three or more options to treat a patient: surgery, drug or no treatment (Imbens, 2000). Similarly, we are interested in the following questions: Did PS-TRE proficiency scores differ significantly among participants of None, Non-formal, and Formal AET programs?

Under Rubin's causal model, where there are multiple treatment conditions available, each individual is associated with a vector of potential outcomes with one value for each condition. Therefore, for T treatment conditions, each individual will have T generalized propensity scores. Imbens (2000) defined the generalized propensity score (GPS) as the conditional probability of each individual ***i*** receiving treatment condition z given observed covariates X:

$$GPS = Pr(Z_i = z|X) \qquad \textbf{Equation 8}$$

To estimate GPS for multiple treatment conditions, multinomial logistic regression models are used. In this study, the following baseline-category logit model (Agresti, 2002) is used to estimate the probability of full-time professionals participating in one format of AET programs instead of the reference category (i.e., None), for Non-formal and Formal AET participation respectively:

$$log \frac{Pr((Z_i = z \mid X)}{Pr((Z_i = Z \mid X)} = \alpha_z + \beta_z X \qquad \textbf{Equation 9}$$

where $Pr(Z_i = z| X)$ is the probability of condition z, $Pr(Z_i = Z \mid X)$ is the probability of the reference category Z, and β_z represents the effect of covariate X on the logit of receiving treatment condition z instead of the reference treatment condition Z (Leite, 2017). This model, in terms of the probability of selecting each treatment condition, is as follows (Agresti, 2002):

$$Pr(Z_i = z \mid X) = \frac{exp(\alpha_z + \beta_z X)}{1 + \sum_{z=1}^{C-1} exp\,(\alpha_z + \beta_z X)}, \quad z = 1, \dots C\text{-}1 \qquad \textbf{Equation 10}$$

$$Pr(Z_i = Z \mid X) = \frac{1}{1 + \sum_{z=1}^{C-1} exp\,(\alpha_z + \beta_z X)} \qquad \textbf{Equation 11}$$

Where $\boldsymbol{C}$ is the total number of conditions, with $\boldsymbol{\alpha}_z = 0$ and $\boldsymbol{\beta}_z = 0$ for the reference category $\boldsymbol{Z}$. The probabilities of receiving exactly one of the treatment conditions (including the reference category) should sum to 1 for each individual $\boldsymbol{i}$.

Each GPS indicates the probability of being in a specific treatment condition versus being in the reference category. The generalized propensity scores for C-1 treatment conditions are estimated using Equation 10 while the generalized propensity score for the reference category is estimated using Equation 11.

Table 9 lists observed pre-treatment covariates that will be used to fit the propensity score model for each format of AET programs. Note that this list is almost a reproduction of predictors for full-time professionals' PS-TRE proficiency scores (see Table 6 for a complete list of predictors selected within the theoretical framework for factors related to PS-TRE). Instead of proficiency scores in PS-TRE, full-time professionals' probabilities of participating in the different formats of AET programs (i.e., propensity scores) are estimated based on explanatory variables listed below.

Table 9. Observed Pre-Treatment Covariates for Propensity Score Estimation

<u>Clusters</u>	<u>*Explanatory Variables*</u>
1) Socio-Demographic Background	*Age (in 10-year bands), Gender, Family Background (proxied by Parental Education), Educational Attainment*
2) Occupational Categories	*Science & Engineering, Health, Teaching, Business & Administration, Information & Communications Technology (ICT) Legal, Social & Cultural*
3.1) Use of Key Information-Processing Skills at Work	*Reading, Writing, Numeracy, ICT*
3.2) Use of Key Information-Processing Skills at Home	*Reading, Writing, Numeracy, ICT*
4) Use of Generic Workplace Skills / Job-Related Activities	*Influencing, Planning, Task Discretion, Learning at work*

Outcome Variable:

Full-time Professionals' Probabilities of Participating in the Different Formats of AET Programs

3.3.2 Assessment of Area of Common Support

Once propensity scores are estimated, it is important to evaluate the success of the process. According to Ho and colleagues, the estimation of propensity scores can be considered successful if "in combination with a matching, stratification, or weighting strategy, they are able to produce adequate balance of covariate distributions between treated and untreated samples" (Ho, Imai, King & Stuart, 2007). Specifically, there are two diagnostic measures of propensity score estimation: (1) whether the estimation method converged and produced estimates between

0 and 1; (2) whether there is an adequate area of common support “which is the region of the distribution of propensity scores where value exist for both treated and untreated cases” (Leite, 2017).

When there are multiple treatment conditions, common support should hold for all pairs of treatment conditions with respect to each vector of GPS. That is to say, for each vector of GPS and each pair of treatments being compared, there should be an adequate “range of the generalized propensity score distribution of the treatments where there are individuals with similar probabilities of receiving each treatment” (Leite, 2017).

3.3.3 Evaluation of Covariate Balance

As the main indicator of the success of the propensity score method, covariate balance provides evidence that the distribution of each covariate for treated and untreated groups is similar. Therefore, bias due to these observed covariates has been removed. In practice, there are three measures that serve to quantify covariate balance:

- Standardized difference between the weighted means of the treated and untreated groups. Mean difference can be standardized with pooled standard deviations or the standard deviation of one of the groups (Stuart & Rubin, 2007; Stuart, 2010; Austin, 2011).
- Ratio of the residual variances of the treated and untreated groups after adjusting for the propensity score. The residual variance for each covariate is obtained by regressing the covariate on the propensity score, for treated and untreated groups, respectively (Stuart & Rubin, 2007; Stuart, 2010).
- The mean and maximum differences between covariate distributions in empirical QQ-plots (Ho, Imai, King & Stuart, 2007).

When there are multiple treatment conditions, the assessment of covariate balance can be performed between all possible pairs of conditions. The absolute values of the standardized effect sizes between pairs of conditions for all covariates are calculated and summarized across all treatment conditions. Generally, standardized mean differences of less than .20 are considered small, .40 are considered moderate, and .60 are considered large (Cohen, 1988). If the covariate balance criterion is not achieved with some covariates, three remedial actions can be taken: re-specifying the propensity score model, changing the propensity score estimation method, or including the covariates that did not reach the specified criteria for covariate balance in the outcome model (Leite, 2017).

3.3.4 Estimation of Treatment Effect

Once covariate balance is achieved, estimation of treatment effect can be performed. Depending on the propensity score method of choice and the nature of the outcome, a variety of parametric or non-parametric estimations are available, such as weighted mean differences, weighted regression and regression-adjusted weighted mean differences (Imbens, 2004; Lunceford & Davidian, 2004; Schafer & Kang, 2008), as well as with complex statistical models, such as multilevel models (Leite et al., 2015) and structural equation models (Leite, Sandbach, Jin, MacInnes, & Jackman, 2012).

Treatment effects of the different formats of AET programs on full-time professionals' PS-TRE proficiency scores can be obtained by comparing all possible pairs of conditions – Non-formal vs. None, Formal vs. None, and Non-formal vs. Formal. The difference from each pairwise comparison estimates the average treatment effect (ATE) of one particular treatment relative to another. Table 10 describes causal estimands for the effectiveness of the different formats of AET programs on full-time professionals' PS-TRE proficiency scores.

Table 10. Estimates of Average Treatment Effects (ATEs) for Three Treatment Levels

Pairwise Comparisons	Pairwise ATEs	Interpretations
Non-formal vs. None	$\boldsymbol{ATE_1}$ - $\boldsymbol{ATE_0}$ = PS-TRE$_{t1}$ - PS-TRE$_{t0}$	Effectiveness of Non-formal AET programs (vs. None)
Formal vs. None	$\boldsymbol{ATE_2}$ – $\boldsymbol{ATE_0}$ = PS-TRE$_{t2}$ - PS-TRE$_{t0}$	Effectiveness of Formal AET programs (vs. None)
Non-formal vs. Formal	$\boldsymbol{ATE_2}$ – $\boldsymbol{ATE_1}$ = PS-TRE$_{t2}$ - PS-TRE$_{t1}$	Effectiveness of Non-formal AET programs (vs. Formal)

Pairwise weighted mean differences are commonly used to estimate the ATE of one particular treatment condition relative to another:

$$\Delta_{21} = \frac{\sum_{i=1}^{n_2} W_{i2} Y_{i2}}{\sum_{i=1}^{n_2} W_{i2}} - \frac{\sum_{i=1}^{n_1} W_{i1} Y_{i1}}{\sum_{i=1}^{n_1} W_{i1}} \qquad \textbf{Equation 12}$$

On the left side of the formula, $\boldsymbol{\Delta_{21}}$ is the ATE of treatment condition 2 compared with treatment condition 1. On the right side, $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ are the number of participants in each treatment condition and $\boldsymbol{Y_{i1}}$ and $\boldsymbol{Y_{i2}}$ are the outcomes.

For individuals with higher propensity scores, selection on the observed covariates plays a stronger role in their choice of treatment. To obtain unbiased estimates of the treatment effect, we use the reciprocal of propensity scores to "downplay" the contributions of individuals whose choices are more predictable, in comparison to the contribution of those whose choice is less so (Murnane & Willett, 2011, p. 326). Specifically, the inverse of the propensity score will be used as the propensity score weight "which balance the pre-treatment characteristics between the group and the entire population" (McCaffrey et al., 2014). As per McCaffrey et al.'s tutorial, a consistent estimate of the ATE (i.e., the population mean of potential outcomes from a treatment of interest) is given by the weighted mean:

$$\frac{\sum_{i=1}^{n} \boldsymbol{T_i}\,[\boldsymbol{t}]\,\boldsymbol{Y_{i}}\boldsymbol{W_i}[\boldsymbol{t}]}{\sum_{i=1}^{n} \boldsymbol{T_i}\,[\boldsymbol{t}]\boldsymbol{W_i}[\boldsymbol{t}]}$$

Equation 13

Where the weight $\boldsymbol{W_i}[\boldsymbol{t}]$ is the inverse of the propensity score or the inverse of the probability that an individual $\boldsymbol{i}$ with pre-treatment characteristics $\boldsymbol{X}$ receives treatment $\boldsymbol{i}$. Treatment effects of interest can be estimated by applying Equation 13 for the estimands in Table 10 "with the unknown population means replaced by their estimates" (McCaffrey et al., 2014). For instance, we estimate the ATE of Formal AET programs relative to None, $\boldsymbol{ATE_2}$ - $\boldsymbol{ATE_0}$ = $PSTRE_{t2} - PSTRE_{t0}$, by the $\widehat{\boldsymbol{ATE_2}}$ - $\widehat{\boldsymbol{ATE_0}}$ = $\widehat{\boldsymbol{PSTRE}}_{\boldsymbol{t2}}$ - $\widehat{\boldsymbol{PSTRE}}_{\boldsymbol{t0}}$ and similar replacements would be used to estimate the ATEs of other two formats of AET programs. It has been supported by empirical evidence that the application of inverse probability-of-treatment weighting (IPTW) in estimating treatment effects allows for more robust inferences when faced with more than two treatment groups (McCaffrey et al., 2014).

RQ4. To what extent are the estimated associations between full-time professionals' PS-TRE proficiency scores and their participation in the different formats of AET programs robust to hidden bias due to unobserved confounders?

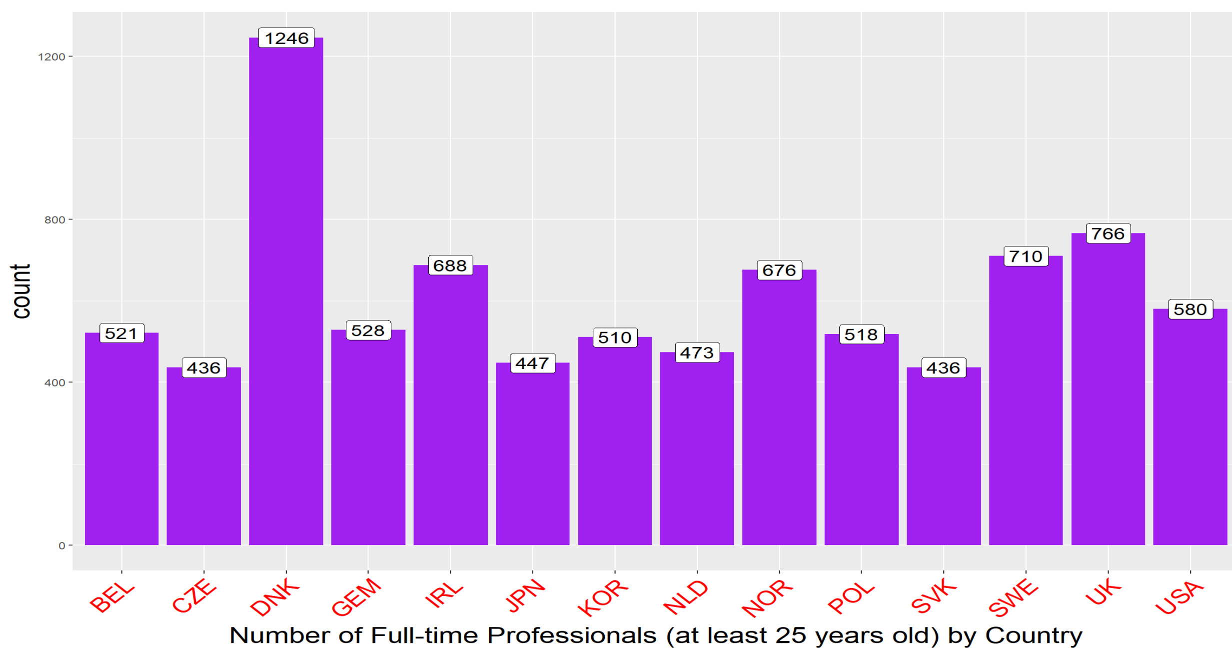
To check the robustness of the estimated associations between full-time professionals' PS-TRE proficiency scores and their participation in the different formats of AET programs, two sets of parallel analyses are conducted – one with a similar but different sample and the other with two alternative outcomes that are supposed to be highly correlated with PS-TRE proficiency scores. More specifically, we first estimate the associations between PS-TRE proficiency scores and participation in the different formats of AET programs for full-time associates in the 14 selected countries and compare the results with that for full-time professionals. Among all other groups of the PIAAC population (e.g., full-time employed, part-time employed, unemployed, student, permanently disabled, etc.), full-time associates are the closest to full-time professionals not only in terms of employment status but also in terms of skill sets. If the results from the sample of full-time professionals share similar patterns with that from the alternative sample of full-time associates, our confidence in the robustness of the estimators to the presence of hidden bias (due to unobserved confounders that are not correlated with the observed covariates but are correlated with the outcome) can be greatly boosted. In a like manner, we estimate the associations between full-time professionals' Literacy and Numeracy proficiency scores and their participation in the different formats of AET programs. Given the strong correlation between PS-TRE proficiency and Literacy and Numeracy proficiency (OECD, 2013c), it is natural to expect that the patterns observed in the main analyses (i.e., RQ3) will retain in the parallel analyses where Literacy and Numeracy proficiency scores are the outcome variables (i.e., RQ4).

3.4 Description of Full-time Professionals Sample in the 14 Selected Countries

As described in Section 3.1, the final sample of this study contains 8,535 full-time professionals from the 14 selected countries who were at least 25 years old at the time of the survey. Figure 1 shows the sample size is the largest in Denmark (N = 1,246), 1.6 times greater than the sample size of the second-largest country (766 in the United Kingdom). The Czech Republic and the Slovak Republic have the smallest number of full-time professionals (436 in each), closely followed by Japan (447). Such cross-country variations are in accordance with what was found in Table 5. On average across the 14 selected countries, full-time professionals aged 25 and above comprise just 9.9% of the entire PIAAC sample (before applying sampling weights) – the percentages are the highest among the three Scandinavian countries -- 17% in Denmark, 15.9% in Sweden and 13.2% in Norway -- and the lowest in the so-called Visegrád

Group countries -- 5.5% in Poland, 7.1% in the Czech Republic and 7.6 in the Slovak Republic. What is more, the United States also has a high percentage (11.6%) of the sampled respondents working as full-time professionals, second only to those Scandinavian countries. Lastly, although the United Kingdom has the second-largest number of full-time professionals in the final sample, its percentage of full-time professionals (8.6%) is among the lowest across countries – only slightly higher than the Visegrád Group countries and the two East Asian countries (i.e., South Korea and Japan).

Figure 1. The Distribution of Full-time Professionals (≥ 25 years old) by Country



3.4.1 Socio-Demographic Background Factors and Occupational Categories of Sampled Full-time Professionals in the 14 Selected Countries

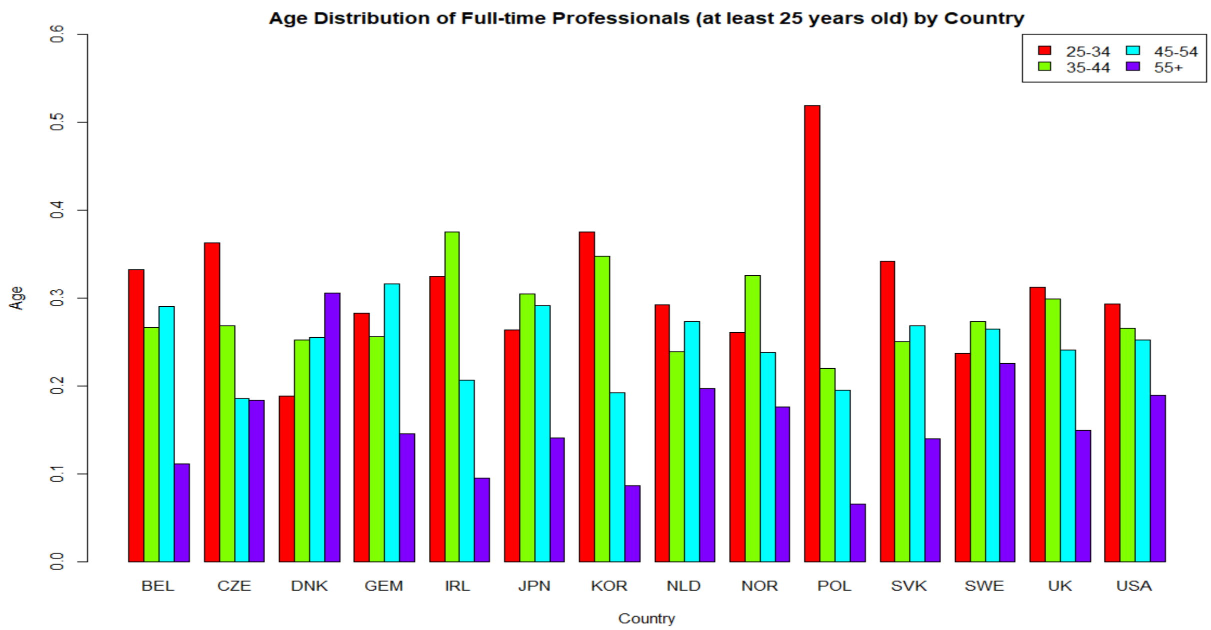
Several socio-demographic factors from the Background Questionnaire (BQ) were analyzed to describe the sample at the individual country level, including: age cohort (in 10-year bands), gender, highest level of educational attainment, highest level of parental education as well as the occupational categories that full-time professionals are in.

Age (in 10-year bands):

On average across the 14 selected countries, the oldest cohort (i.e., 55-65 year olds) comprises less than one fifth (17.1%) of full-time professionals; while the youngest cohort (i.e., 25-34 year olds) comprises almost one third (29.9%), closely followed by the 35-44 and the 45-54 age groups (28.2% and 24.8% respectively). In other words, the proportion of full-time professionals reaches a peak between the young ages of 25 and 34 and declines steadily, with the oldest cohort representing a smaller share of this particular labor market than the younger cohorts.

As shown in Figure 2, full-time professionals aged 55 and above constitute the smallest fraction of the total of full-time professionals in all but one country (i.e., Denmark), but the percentage ranges from a low of 6.6% in Poland to a high of 18.3% in the Czech Republic. On the other hand, the so-called “workers of prime age, i.e. aged 25-54” (OECD, 2013c) remain the mainstay of full-time professionals in these countries. In particular, workers at a younger age (i.e., 25-34 year olds) are taking over and have become the largest proportion of full-time professionals in eight countries. In Poland, more than half (51.9%) of full-time professionals are below the age of 35, followed by South Korea (37.5%), and the other two Visegrád Group countries (36.2% in the Czech Republic and 34.2% in the Slovak Republic).

Figure 2. Age Distribution of Full-time Professionals (≥ 25 years old) by Country



Denmark is distinct from all other countries in the age distribution of full-time professionals (at least 25 years old). Contrary to the common theme that prime-age workers are the major component of full-time professionals, the oldest cohort is still going strong in Denmark, comprising almost one third (30.5%) of Danish full-time professionals. While the oldest cohort (i.e., 55-65 year olds) makes up the largest proportion of full-time professionals in Denmark, the youngest cohort (i.e. 25-34 year olds) only accounts for as little as 18.8% -- smaller than any age cohort in Denmark. In fact, Denmark is the country where the smallest proportion of 25-34-year-old full-time professionals was observed.

Gender:

On average, the gender composition of full-time professionals across the 14 selected countries (i.e., 49.3% female vs. 50.7% male) is in line with the belief that "many OECD countries have made significant progress over the past few decades in narrowing the gender gap in education and employment". (OECD, 2013c). However, further examinations of the gender distribution of full-time professionals (at least 25 years old) at the individual country level conveyed mixed messages. As evident in Figure 3, all the three Visegrád Group countries and the United States have witnessed a much higher percentage of women than men working as full-time professionals. In Poland, as high as 62.9% of full-time professionals are female, followed by the Slovak Republic (58.9%) and the Czech Republic (57.6%). The United States also boasts 54.7% of full-time professionals identified as female. What is more, a more or less gender-balanced pool of full-time professionals was established in all the three Scandinavian countries and the United Kingdom. On the other hand, a few countries are still dealing with a startling gender ratio in favor of men. In the Netherlands, male professionals are about 2.6 times more numerous than female professionals (71.9% vs. 28.1%, or delta = 43.8%). In Germany, there are 23.4% more male professionals than female professionals (61.7% vs. 38.3%), followed by the two East Asian countries -- 15% more in South Korea (57.5% vs. 42.5%) and 13.2% more in Japan (56.6% vs. 43.4%).

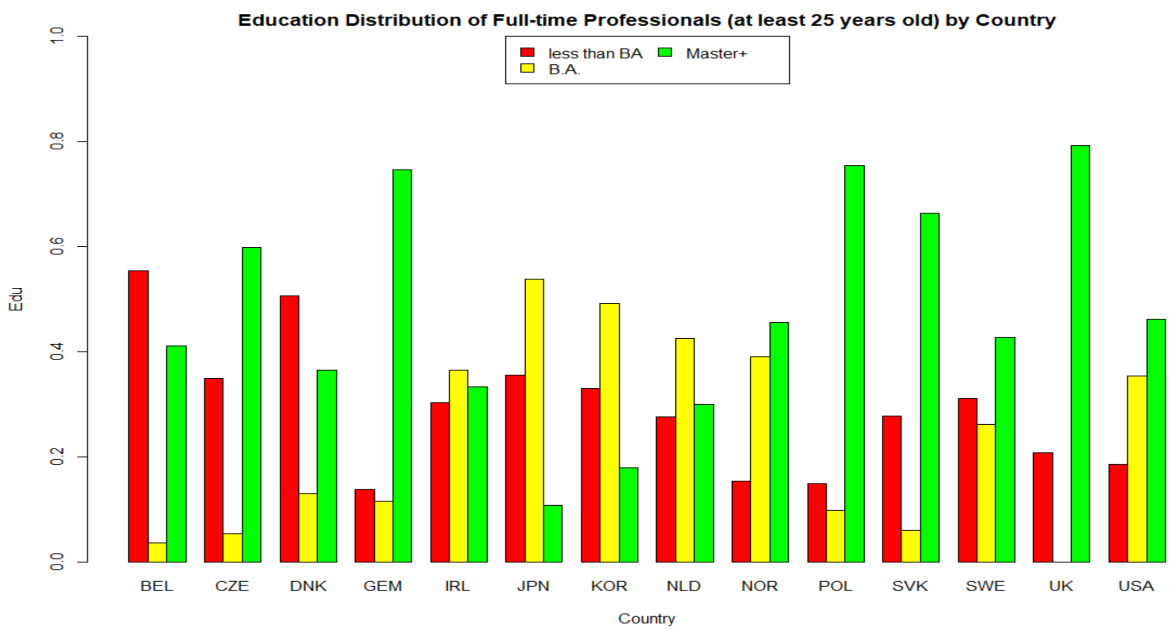
Figure 3. Gender Distribution of Full-time Professional (≥ 25 years old) by Country

Highest Level of Educational Attainment:

In the final sample of full-time professionals (at least 25 years old) from the 14 selected countries, 22.7% of survey respondents reported that they have a Bachelor's degree only and as high as 46.8% reported having a Master's degree or above. As shown in Figure 4, the two East Asian countries enjoy the highest proportion of full-time professionals who have a Bachelor's degree only (53.7% in Japan and 49.2% in South Korea). However, they also have the lowest proportion of full-time professionals with a Master's degree or above (10.7% in Japan and 17.8% in South Korea). On the other hand, countries in Central Europe, including Germany and the three Visegrád Group countries, boast the largest proportion of full-time professionals who hold at least a Master's degree, the percentage ranges from about 75% in Germany and Poland to 66%

in the Slovak Republic to 60% in the Czech Republic[23]. Interestingly, they also have the second-smallest proportion of full-time professionals with a Bachelor's degree only, the percentage ranges from about 5% in the Czech Republic and the Slovak Republic to 9.85% and 11.55% in Poland and Germany. In other words, it is striking to notice how extreme the ratios of Master+ level to Bachelor level full-time professionals are in these Central European countries; that is 6.5 to 1 in Germany, 7.6 to 1 in Poland and 11 to 1 in the other two Visegrád Group countries.

Figure 4. Education Distribution of Full-time Professionals (≥ 25 years old) by Country



Belgium is very similar to the Central European countries because of (1) the smallest proportion of Bachelor's degree holders and (2) the extremely high Master+ to Bachelor ratio among Belgian full-time professionals (up to 11 to 1). However, unlike the Central European countries, Belgium has the largest proportion (55.3%) of full-time professionals who have not

[23] Arguably, the United Kingdom has the highest percentage (79.2%) of full-time professionals with a Master's degree or above; however, our sample shows that no full-time professional in the UK reported having a Bachelor's degree only. For the sake of fairness, the British sample was not to be compared with the other 13 countries.

completed a college education, followed by 50.6% in Denmark. By contrast, Germany, Poland and Norway are among countries with the smallest proportions (between 14% and 15%) of full-time professionals without a Bachelor's degree.

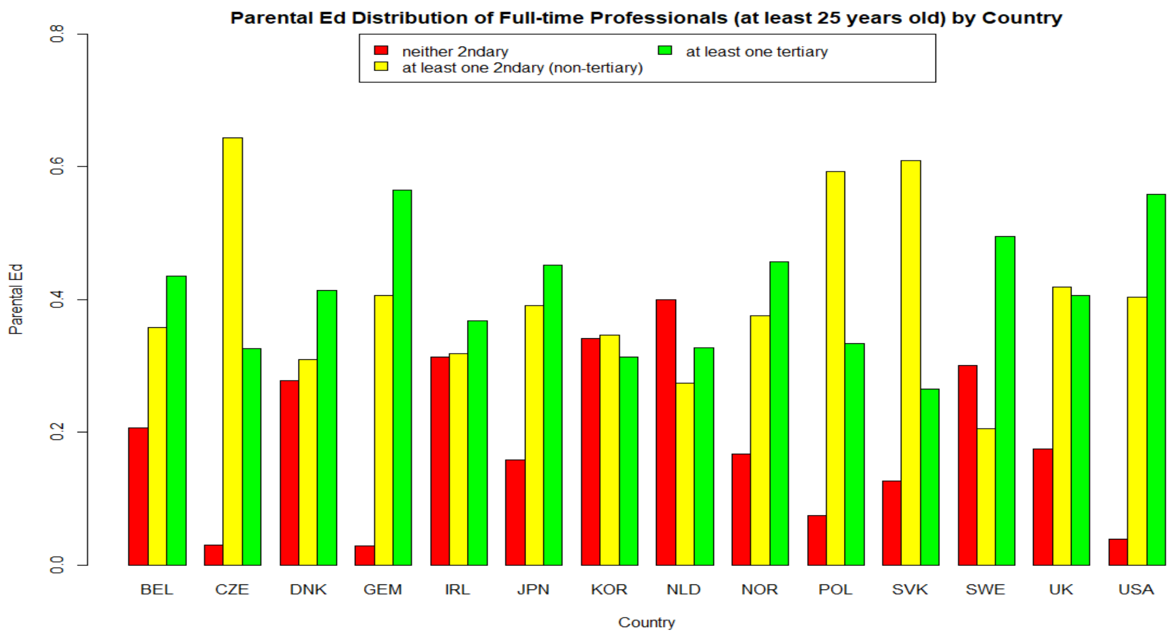
Highest Level of Parental Education:

As per the Survey of Adult Skills (OECD, 2013c), parents' educational attainment offers not only a way to characterize socio-economic background, but also "insights into intergenerational social mobility". In the sample of full-time professionals, an average of 41.4% respondents reported having at least one parent who has attained tertiary education and 38.6% reported having at least one parent who has attained secondary or post-secondary, non-tertiary education. Only one out of every five full-time professionals has neither parent with a high school (upper secondary) education. Figure 5 provides a more granular view of the distribution of parental education levels within and across countries. Germany and the United States boast the highest percentage of parents with tertiary education and the lowest percentage of parents without upper secondary education -- 56.5% vs. 2.9% and 55.8% vs. 3.8% respectively. Given the fact that the German sample contains the largest proportion of highly educated respondents, it is not so surprising to see Germany stands out as the country with the highest proportion of parents who have completed tertiary education. The high proportion of tertiary-educated parents in the American sample, on the other hand, comes a bit unexpected. As you may recall back in Figure 4, rather than having a large proportion of full-time professionals with either a Bachelor's or a Master's degree, the United States only ranked in the middle of the 14 selected countries with a proportion slightly above average. To put this in perspective, it seems that the United States may have lost ground to countries in Central Europe and East Asia with regard to the intergenerational mobility in education. Unlike their parents' generations, the educational attainment of U.S. full-time professionals in the sample does not compare favorably with that of their international competitors.

In addition to Germany and the United States, the three Scandinavian countries and Japan all have relatively high proportions of parents with tertiary education. However, the proportion of "neither parent attained upper secondary education" is much higher in these countries than in Germany and the U.S., ranging from 15.8% in Japan to 30% in Sweden. On the other hand, the highest percentages of non-tertiary-educated parents were observed in the Netherlands (39.9%), South Korea (34.1%) and Ireland (31.4%). Note that this percentage is twice as large in South Korea as that in Japan (i.e., 34.1% vs. 15.8%). Nonetheless, the share of full-time professionals with at least a Bachelor's degree is roughly equal in both countries, indicating a larger degree of intergenerational educational mobility in South Korea than in Japan.

What is more, the Visegrád Group countries boast the highest percentages of parents with upper secondary (non-tertiary) education and the lowest percentages of parents without upper secondary education – 64.3% vs. 3% in the Czech Republic, 59.3% vs. 7.4% in Poland and 60.9% vs. 12.6% in the Slovak Republic respectively. Given the fact that these countries boast the largest proportion of Master+ level but the second-smallest proportion of Bachelor level full-time professionals (see Figure 5), an upward intergenerational mobility in education is evident within the Visegrád Group.

Figure 5. Parental Ed. Distribution of Full-time Professionals (≥ 25 years old) by Country



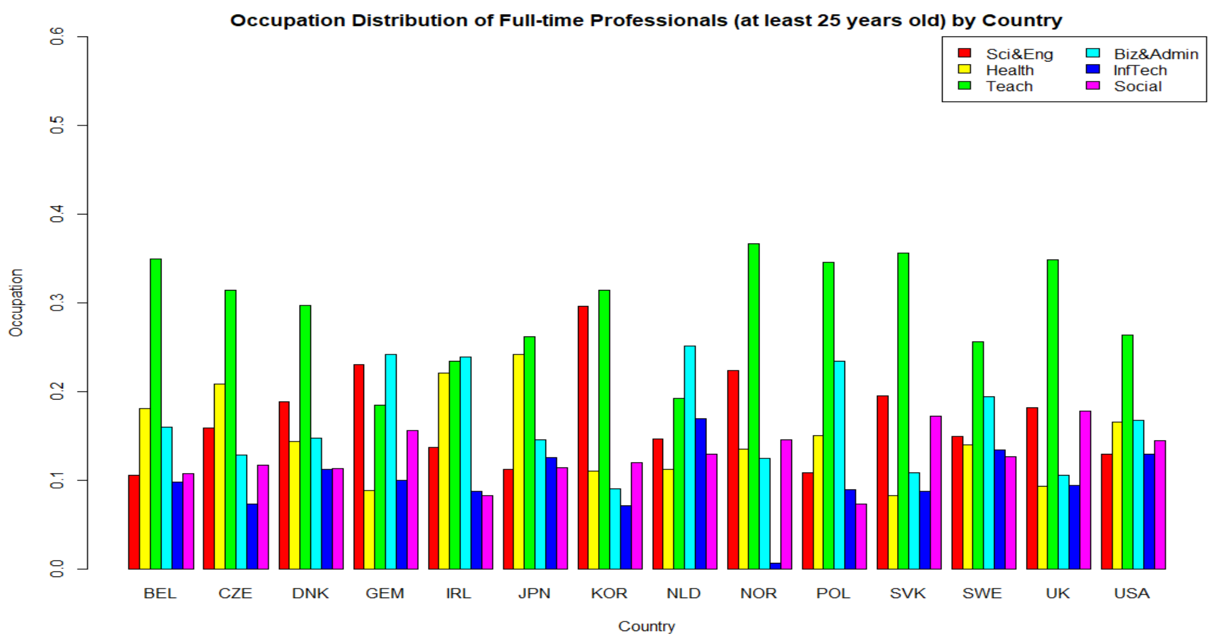
Occupational Categories for Full-time Professionals:

Based on the sample of 8,535 full-time professionals from the 14 selected countries, Figure 6 displays the distribution of the six occupational categories within and across countries. It is clearly shown that Teaching is the most dominant occupation among full-time professionals in all but three countries. On average across countries, almost one third of (29.3%) the sampled respondents are teaching professionals, followed by professionals in Science & Engineering and Business and Administration (about 17% in each category).

Norway, Belgium, the United Kingdom, as well as the three Visegrád Group countries all have relatively large proportions of full-time professionals working in the teaching field, ranging from 36.7% in Norway to 31.4% in the Czech Republic. Even countries with the smallest proportion of teaching professionals (i.e., the Netherlands and Germany) have nearly one fifth of the sampled respondents belonging to this occupational category.

In the other three countries (the Netherlands, Germany and Ireland) where Teaching is not the most dominant occupation, Business & Administration emerges as the largest occupational category, accounting for about one quarter of full-time professionals in each country. However, it is important to know that teaching professionals still make up a relatively large part of the sampled respondents even in these countries, the percentage ranges from a low of 18.4% in Germany to a high of 23.4% in Ireland.

Figure 6. Occupation Distribution of Full-time Professionals (≥ 25 years old) by Country



Science & Engineering professionals are another key component of full-time professionals. Among the 14 selected countries, South Korea has the highest proportion (30%) of full-time professionals working in the Science & Engineering field, which is only 1% less than the largest occupation (i.e. Teaching). What is more, full-time professionals from these two occupational categories represent almost two thirds (61%) of the South Korean sample. This pattern was also observed in Norway, the Slovak Republic and the United Kingdom, where the combination of teaching and Science & Engineering professionals comprises the majority of the sampled respondents, ranging from 59% to Norway and 53% in the United Kingdom.

Following South Korea, Germany has the second highest proportion (23%) of Science & Engineering professionals, which is, again, only 1% less than the largest occupation (in this case, Business & Administration). Together, the two occupations represent almost half (47%) of the German sample. On the other hand, the lowest proportion of Science & Engineering professionals was observed in Belgium and Poland (about 10%) which is less than one third of the size of teaching professionals in both countries.

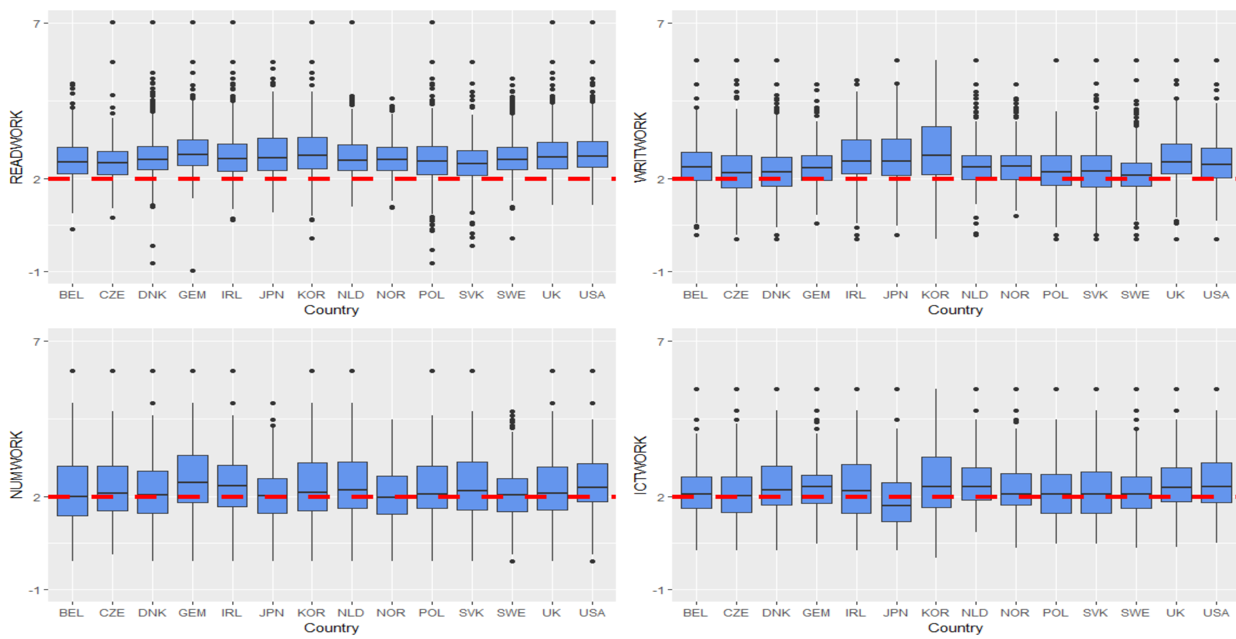
3.4.2 Skills Use at Work, at Home and Job-Related Activities of Sampled Full-time Professionals in the 14 Selected Countries

In addition to socio-demographic factors, the background questionnaire also asked how different skills are used in the workplace. As introduced in Section 3.2, eight indicators of skills use at work were retained for this analysis, four of which correspond to key information-processing skills (reading, writing, Numeracy and ICT skills) and the remaining four refer to generic workplace skills/job-related activities (influencing, planning, task discretion and learning at work). To facilitate comparisons between not only the use of different skills but also the use of the same skill in different settings, four extra indicators of skills use at home were added for the aforementioned key information-processing skills. According to the OECD documents (OECD, 2013b, 2013c), all indicators have been standardized to have a grand mean equal to 2 and standard deviation equal to 1 across the entire PIAAC sample in all 23 participating countries (appropriately weighted).

Use of Key Information-Processing Skills at Work

Figure 7 demonstrates how the use of key information-processing skills at work compare across the 14 selected countries. As it turns out, full-time professionals in the sample reported Reading and Writing as the most frequently used key information-processing skills at work. Overall, the average use of Reading and Writing skills at work among full-time professionals is 2.7 and 2.4 respectively, both are higher than the standardized mean of 2 (and standard deviation of 1) across the pooled sample of respondents in all 23 participating countries with appropriate weights (OECD, 2013b, 2013c). For each country, the average use of both Literacy skills at work is well above the grand mean. Reading skills are reported to be used at work most frequently in South Korea and Germany (> .85 standard deviations above 2) and least in the three Visegrád Group countries (< .6 standard deviations above 2). Similarly, Writing skills are used most frequently in South Korea, Ireland and the United Kingdom (> .65 standard deviations above 2) and least in two of the Scandinavian countries (2.16 and 2.27 in Sweden and Denmark respectively). What is more, full-time professionals in the three Visegrád Group countries also reported relatively low frequency of Writing skills use at work (2.28). In short, among the sampled full-time professionals, both Literacy skills are used most intensively at work in South Korea and least in the three Visegrád Group countries.

Figure 7. Distribution of the Use of Key Information-Processing Skills at Work



On average across the 14 selected countries, full-time professionals reported using Numeracy skills at work as frequently as the entire PIAAC sample. In fact, the average use of Numeracy skills at work is greater than or equal to the grand mean in all the 14 selected countries. Numeracy skills are reported to be used at work most frequently in Germany and the United States (2.53 and 2.41 respectively), followed by Ireland and the Netherlands (> .30 standard deviations above 2). Norway is the country whose average use of Numeracy skills at work is close to the grand mean (2.02), followed by Japan and Sweden (2.06 and 2.07 respectively). You may also notice that in spite of its high frequencies of Literacy skills use at work, South Korea only lands in the middle of the pack when it comes to Numeracy skills (exactly .30 standard deviations above 2).

The use of ICT skills at work follows a somewhat similar pattern to that of Literacy skills. The average use is 2.3 across countries and all but one country (i.e. Japan) reported using

ICT skills at work with above-the-grand-mean frequency. Among the sampled full-time professionals, ICT skills are used most frequently at work in South Korea and the United States (> .45 standard deviations above 2) and least in Japan (1.87). Moreover, full-time professionals in the three Visegrád Group countries also reported relatively low frequency of ICT skills use at work, ranging from 2.10 to the Czech Republic to 2.16 in Poland.

To summarize, for full-time professionals, there are three major findings from comparing the use of key information-processing skills at work across the 14 selected countries:

(1) The two East Asian countries are quite different in that South Korea is the country that ranked consistently at the top in the average use of Literacy and ICT skills in the workplace, while Japan ranked mostly near the bottom in the average of Numeracy and ICT skills.

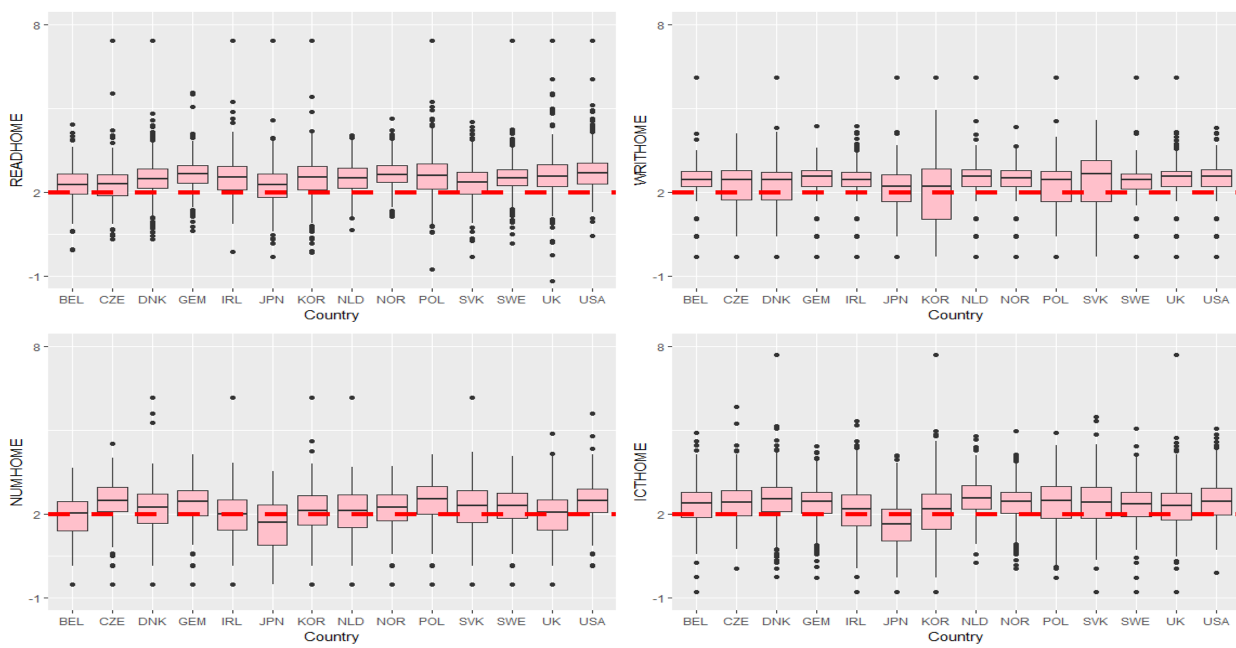
(2) Full-time professionals in the three Visegrád Group countries did not report using Literacy and ICT skills at work as intensively as their international peers.

(3) The use of Numeracy skills at work seemed to have a different distribution among full-time professionals in the 14 selected countries than that of Literacy and ICT skills.

Use of Key Information-Processing Skills at Home

Just as with Figure 7, Figure 8 demonstrates how the use of key information-processing skills at home compare across the 14 selected countries. Unsurprisingly, Reading and Writing skills are the most frequently used skill at home by the sampled full-time professionals. The overall average use of Reading and Writing skills at home is 2.5 and 2.3 respectively (compared with 2.7 and 2.4 at work). Reading skills are reported to be used at home most frequently in the United States (2.73), followed by Germany and Norway (2.67); while Writing skills are used at home most frequently in the Netherlands (2.51), followed by Germany (2.41). Instead of the three Visegrád Group countries, Japan is the country with the lowest frequent use of Reading skills (2.24) and the second-lowest frequency use of Writing skills (2.02) at home. It is important to note that although full-time professionals in South Korea reported using both Literacy skills the most frequently at work (2.89 and 2.75 respectively), they also reported the least use of Writing skills at home (1.98).

Figure 8. Distribution of the Use of Key Information-Processing Skills at Home



Again, in the sample, full-time professionals reported using Numeracy skills at home as frequently as the entire PIAAC sample. Numeracy skills are reported to be used at home most frequently in the United States, the Czech Republic and Poland (> .40 standard deviations above 2), followed by Germany (> .30 standard deviations above 2). Japan is the country with the lowest average use of Numeracy skills at home (1.65), followed by Belgium, Ireland and the United Kingdom (1.85, 1.91 and 1.94 respectively). South Korea also has a relatively low average use of Numeracy skills at home (2.06) among the 14 selected countries.

Interestingly, the distribution of the use of ICT skills at home is more similar to that of Numeracy than Literacy skills. On average across countries, ICT skills are used less frequently at home than at work (2.0 vs. 2.3). The Netherlands and Denmark have the highest average use of ICT skills at home (> .50 standard deviations above 2), followed by the United States, the Czech Republic and Poland (> .40 standard deviations above 2). Again, Japan is the country with the lowest average use of ICT skills at home (1.63), followed by South Korea (2.14). More interestingly, full-time professionals in the three Visegrád Group countries reported relatively low frequency of ICT skills use at home, whereas the opposite is true with their use of ICT skills at work.

In sum, a comparison between the use of key information-processing skills at work and at home reveals the following findings:

(1). Notwithstanding differences in the use of information-processing skills at work, the two East Asian countries are very similar when it comes to the use of these skills at home. Across the four skill domains, both Japan and South Korea were consistently ranked at the bottom in the average use of skills at home. It is particularly important to point out that among the 14 selected countries, South Korea reportedly used these skills with the highest frequency at work, but the lowest frequency at home.

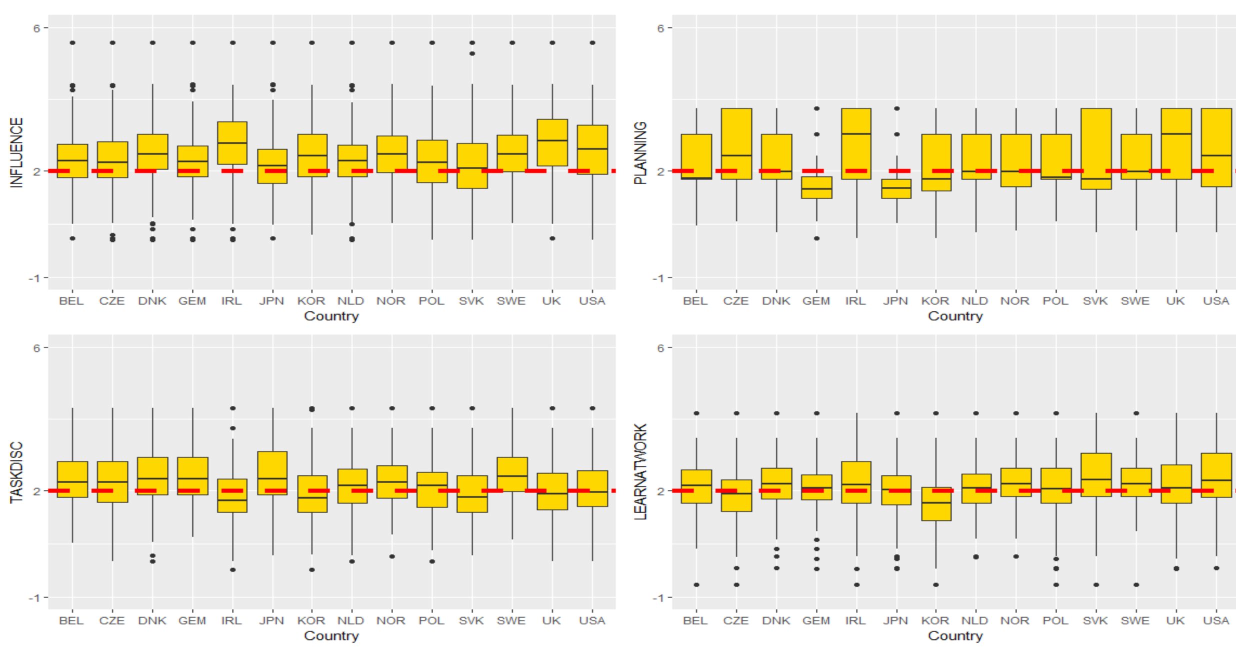
(2). In comparison to full-time professionals in Germany, those in the Netherlands and the United States reported the highest average use of key information-processing skills at home.

(3) Contrary to the previous finding that the Visegrád Group countries have a less intensive use of key information processing-skills at work, the Czech Republic and Poland reported a relatively high average use of Numeracy and ICT skills at home.

Use of Generic Workplace Skills/Job-related Activities

Examination of self-reported use of generic workplace skills/job-related activities delved into different aspects of full-time professionals' competencies at work and revealed different patterns. As evident in Figure 9, the use of generic workplace skills, as well as the rankings of countries, vary substantially. With an overall average use as high as 2.5 and the country mean well above the grand mean in all the 14 selected countries, influencing skills seemed to be the most frequently used generic skills in the workplace. Influencing skills are reported being used the most among full-time professionals in the United Kingdom, Ireland and the United States (> .70 standard deviations above 2) and the least in Japan the Slovak Republic (2.17 and 2.20 respectively).

Figure 9. Distribution of the Use of Workplace Skills/Job-related Activities



The average use of Planning skills across countries is .1 standard deviations less than that of Influencing skill and there are two countries reporting using Planning skills with below-the-grand-mean frequency. The highest average use was also observed in the United Kingdom and

Ireland (> .70 standard deviations above the grand mean) and the lowest is in Japan and Germany (1.67 and 1.72 respectively). A re-examination of Figure 9 shows that the distribution of Planning skills use in the workplace has a wide spread in most countries. The Slovak Republic and the United States are countries with the largest interquartile range (IQR), followed by Ireland and the United Kingdom. In other words, the frequency with which Planning skills are used at work varies greatly in these countries. Also note that Germany and Japan have the smallest IQR across countries and both IQRs are completely underneath the grand mean. That is to say, most full-time professionals in these two countries use Planning skills at work with a frequency below the grand mean.

The distributions of Task Discretion and Learning at Work, on the other hand, are more homogeneous within and across the 14 selected countries. The overall average use of Task Discretion and Learning at Work is 2.3 and 2 respectively. For Task Discretion, the highest average use was observed in Japan (2.5) and the lowest is in Ireland, the Slovak Republic and the United Kingdom (1.89, 1.99 and 2.05 respectively) -- pretty much in reverse order of the country rankings for Influencing and Planning skills (except for the Slovak Republic). As for Learning at Work, the Slovak Republic ascended to the top of the rankings with the average use of 2.42, followed by the United States (2.40) and the three Scandinavian countries (2.31 for each). Just as with the country rankings for Influencing and Planning skills, Japan remains near the bottom for Task Discretion skills and its average use (2.01) is only larger than that of the Czech Republic and South Korea (1.86 and 1.63 respectively).

Skills Used Jointly in the Workplace

The OECD document (2013c) pointed out that (1) "while information-processing skills tend to be used together, generic skills are not" and (2) Influencing skills are the only generic skills that tend to be associated with reading and writing skills. However, this analysis based on full-time professionals only (at least 25 years old) yields different results. The strongest correlation was observed between Influencing and Planning skills in all the 14 selected countries, ranging from .48 in Germany to .62 in South Korea, Poland and the United Kingdom. While the second strongest correlation was observed between Reading and ICT skills at work, ranging from .3 in the United States to .49 in Norway. Note that Influencing skills are still the exception in the sample because they are more closely related with Reading and Writing skills than any other generic workplace skills. Please refer to Appendix E for the overall correlation matrix between all pairs of skills used at work for the sample of full-time professionals.

3.4.3 The Different Formats of AET Participation for Sampled Full-time Professionals in the 14 Selected Countries

The First Results from the Survey of Adult Skills (OECD, 2013c) examined participation in adult education and training (AET) programs among adults aged 25-65 and classified all the participating countries into four groups (p. 209). As shown below in Table 11, only four of the 14 selected countries in the PIAAC sample reported participation rates exceeding 60%. In the sample of full-time professionals aged 25-65, however, all the 14 selected countries exhibit participation rates greater than 64%. When we elevated the cutoff values for different groups and re-ranked the 14 selected countries (see columns 3 and 4 in Table 11). The adoption of heightened standards to the sample only did not affect the original rankings for most countries, except for the United States – restricting the sample to full-time professionals only raised the U.S. ranking of AET participation from the second tier to the very top (89%). That is to say, full-time professionals in the U.S. have a much higher AET participation rate than the general public.

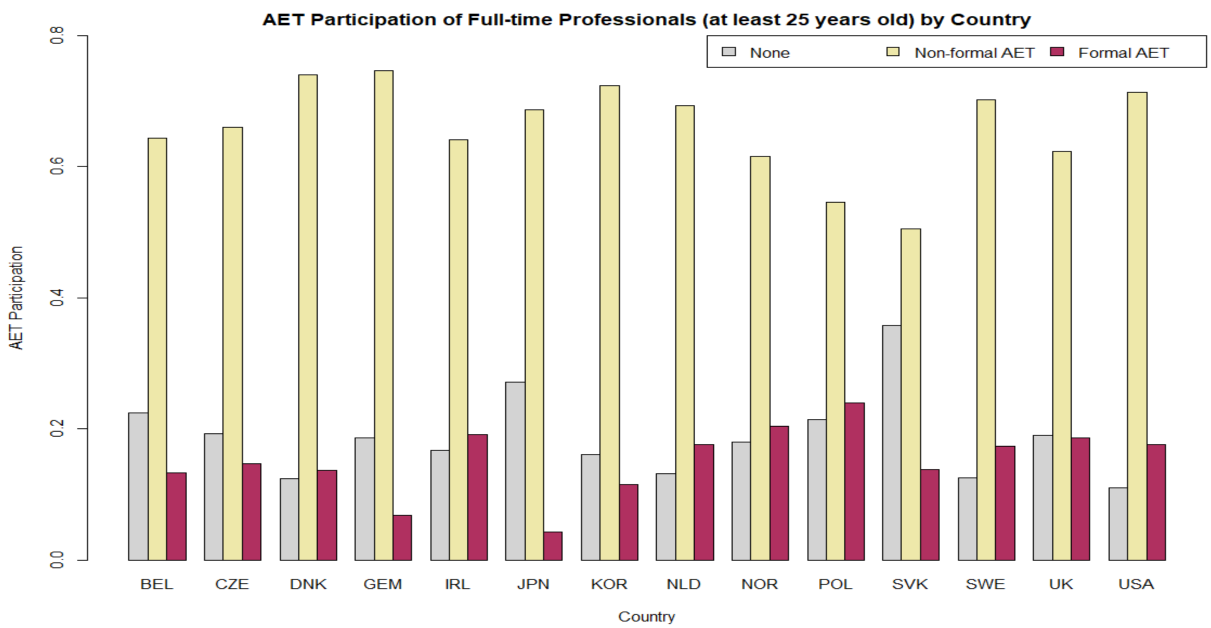
Table 11. Classification of AET Participation Rates, by Country

	Adults aged 25-65		Full-time Professionals aged 25-65	*Full-time Associates Ages 25-65*
Group 1: > 60%	**Denmark, The Netherlands, Sweden,** Norway	New Group 1: > 85%	United States (89%), **Denmark (88%), The Netherlands (87%), Sweden (87%)**	
Group 2: 50-50%	**South Korea, Ireland, United Kingdom, Germany,** United States	New Group 2: 80-84%	**South Korea (84%), Ireland (83%),** Norway (82%), **United Kingdom (81%), Germany (81%)** Czech (81%)	*The Netherlands (81%) Sweden (81%) Norway (80%)*
Group 3: 40-50%	Czech Republic, **Belgium, Japan**	New Group 3: 70-80%	Poland (79%), **Belgium (78%), Japan (73%)**	*United Kingdom (79%) Denmark (77%) Czech (76%) Germany (76%) United States (76%) Ireland (74%) South Korea (71%)*
Group 4: 30-40%	Poland, **Slovak**	New Group 4: 50-70%	**Slovak (64%)**	*Belgium (66%) Poland (63%) Japan (63%) Slovak (58%)*

In addition to the binary indicator (i.e., Yes or No) of full-time professionals' AET participation, Figure 10 delves further into the different formats of AET programs – Formal, Non-forma and None. It is obvious that Non-formal AET programs are the most popular among the sampled full-time professionals. On average across the 14 selected countries, as high as two thirds of full-time professionals reported having participated in Non-formal AET programs during the year prior to the survey. The Non-formal AET participation rate is the highest in Germany (74.6%), followed by Denmark, South Korea and the United States (74%, 72.4% and 71.4%). On the other hand, full-time professionals in the Slovak Republic reported the lowest participation rate for Non-formal AET programs (50.4%), followed by 54.6% in Poland. It is important to note that the Slovak Republic not only has the lowest participation rate for Non-formal AET programs, but also the lowest participation rate for AET programs in general (64%).

In the sample, the proportion of full-time professionals who did not participate in any education or training during the year prior to the survey is slightly higher those who participated in formal AET programs (17.8% vs. 15.5%). As was shown in Figure 10, following the United States (11%), Denmark, Sweden and the Netherlands are countries with the lowest percentages of non-participants (12.4%, 12.5% and 13.1% respectively); while the Slovak Republic, Japan, Belgium and Poland are countries where the highest percentage of non-participants was observed (35.8%, 27.1%, 22.5% and 21.4% respectively). As far as formal AET participation is concerned, more than one fifth (23.9%) of full-time professionals in Poland reported having participated in formal AET programs in the previous 12 months, followed by 20.4% in Norway. On the other hand, the formal AET participation rate is as low as 4.3% in Japan, follow by 6.8% in Germany.

Figure 10. Distribution of AET Participation among Full-time Professionals (≥ 25 years old) by Country

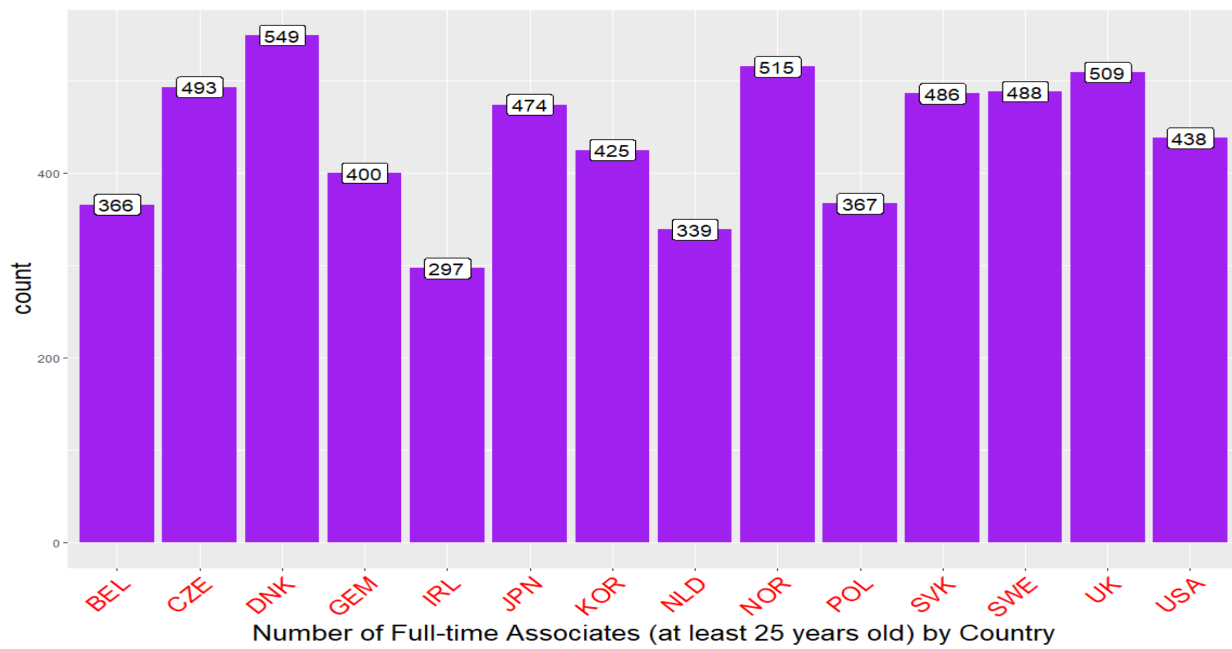


3.5 Description of Full-time Associates Sample in the 14 Selected Countries

Before diving into the detailed results of research questions which involve not only the analysis of full-time professionals but also full-time associated professionals (i.e., full-time associates), it is necessary to get an overview of the sample of full-time associated professionals (short for the associated sample) and make comparisons between the two types of full-time workers. As a reminder, what we have learned from Section 3.4 is that there were 8,535 full-time professionals aged 25 and above from the 14 selected countries, comprising 9.9% of the entire PIAAC sample (before applying sampling weights). The percentages are the highest among the three Scandinavian countries and the lowest in the Visegrád Group countries. For comparison, the associated sample contains 6,146 full-time associates from the 14 selected countries who were at least 25 years old at the time of the survey. Figure 11 (or Figure 1b) shows the sample size ranges from 549 in Denmark to 297 in Ireland. Note that for full-time professionals (see

Figure 1), the Danish sample is 1.6 times greater than the sample of the second-largest country. For full-time associates, on the other hand, the Danish sample is marginally larger than the second-largest country (515 in Norway).

Figure 11. The Distribution of Full-time Associates (⩾ 25 years old) by Country (Figure 1b, corresponding to Figure 1)



Compared with full-time professionals that comprise 9.9% of the entire PIAAC sample, full-time associates aged 25 and above account for an even smaller proportion, i.e., 7.1% (before applying sampling weights). As can be seen in the last column of Table 5, the percentages of full-time associates are the highest in two Scandinavian countries (10.9% in Sweden and 10% in Norway) and the lowest in Poland (3.9%), Ireland (5.0%) and the United Kingdom (5.7%). Note that in Denmark, the Scandinavian country with the highest percentage of full-time professionals (up to 17%), only 7.5% of the sampled respondents self-reported as full-time associates. In fact, the proportion of full-time associates in Denmark is even lower than that in two countries from the Visegrád Group -- 8.1% in the Czech Republic and 8.5% in the Slovak Republic.

Another interesting observation is that the percentages for both types of full-time workers are relatively high in the United States (11.6% for professionals and 8.7% for associates) but relatively low in the United Kingdom (8.6% for professionals and 5.7% for associates).

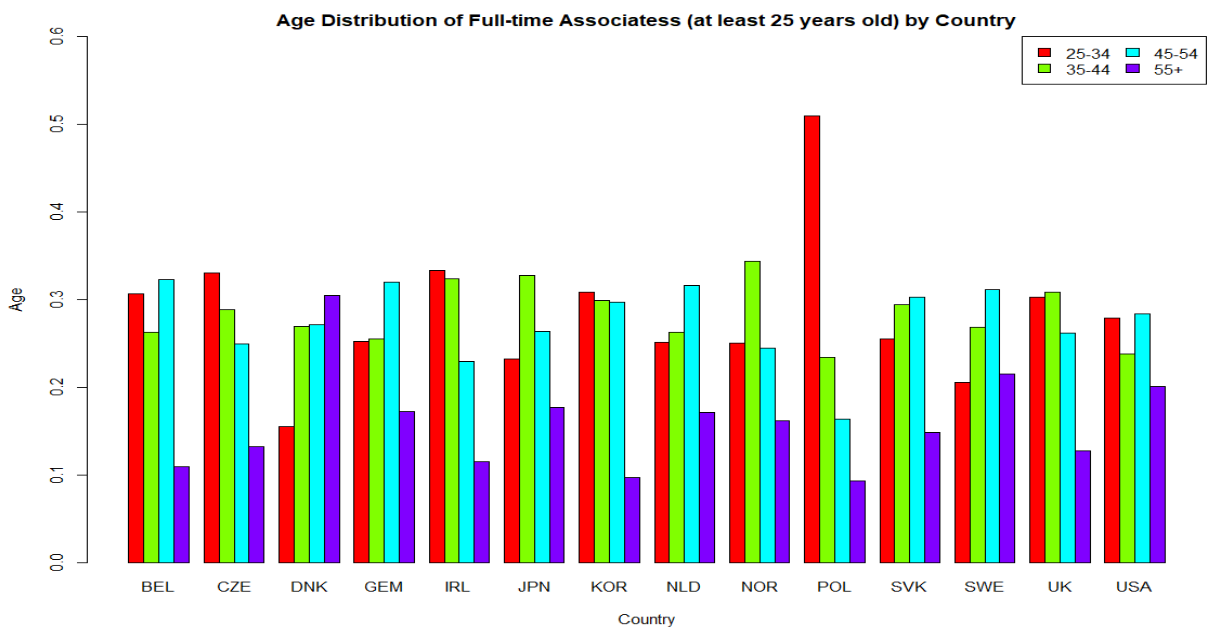
3.5.1. Socio-Demographic Background Factors and Occupational Categories of Sampled Full-time Associates in the 14 Selected Countries

Age (in 10-year bands):

On average across the 14 selected countries, the oldest cohort (i.e., 55-65 year olds) comprises the smallest fraction (16.35%) of the total of full-time associates; while the second-youngest cohort (i.e., 35-44 year olds) makes up the largest proportion (28.52%), closely followed by the younger 25-34 and the older 45-54 age groups (27.69% and 27.43% respectively). Note that unlike full-time professionals, the proportion of full-time associates reaches a peak at older ages (between 35 and 44) and does not start declining until much later (aged 55 and above).

At the individual country level, the age distribution of full-time associates is very similar to that of full-time professionals (see Figure 2 and Figure 12/Figure 2b). The oldest cohort represents the smallest proportion of sampled respondents in all but one country (i.e., Denmark). Poland still has the lowest percentage of 55-65 year olds, that is, 9.3% compared to 23.5% in Sweden and 30.4% in Denmark. On the other hand, “prime-age workers” between 25-54 years-of-age constitute the absolute majority of full-time associates in these countries. In five countries, the youngest cohort (i.e., 24-34 year olds) is taking up the largest proportion of full-time associates – compared with eight countries in the sample of full-time professionals. Again, in Poland, more than half (51%) of full-time associates are below the age of 35, more than double the size of the second largest age group, i.e., 23.4% are between the ages of 35 and 44. As is the case with the sample of full-time professional, the Czech Republic and South Korea are among countries with relatively high percentages of younger full-time associates (33.1% and 30.8% respectively).

Figure 12. Age Distribution of Full-time Associates (⩾ 25 years old) by Country (Figure 2b, corresponding to Figure 2)



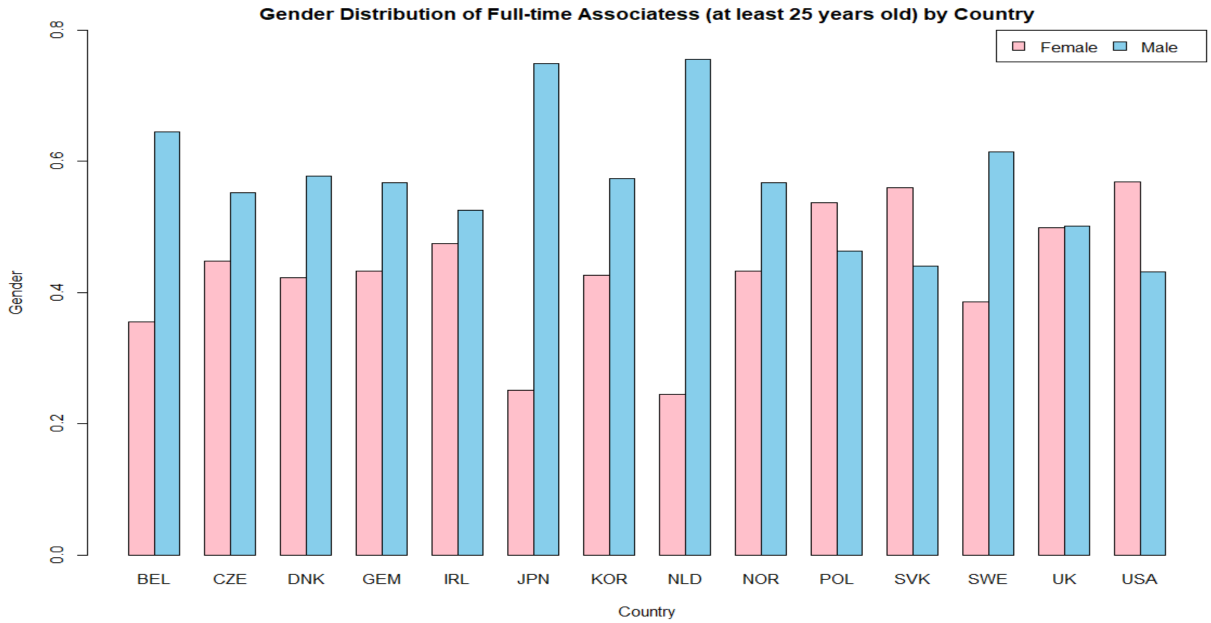
With the associated sample, the age distribution in Denmark once again distinguishes itself from that of all other countries. Similar to the Danish sample of full-time professionals, the oldest cohort makes up the largest proportion (30.4%) of full-time associates while the youngest cohort accounts for as little as 15.5%. Therefore, in both samples, Denmark is the country with the smallest share of full-time workers at a younger age (i.e. 25-34 year olds).

Gender:

Recall that in section 3.4.1., an approximately 1:1 gender composition was observed with the sample of full-time professionals. However, the gender distribution of full-time associates tells a very different story (see Figure 3 and Figure 13/Figure 3b). On average across the 14 selected countries, the percentage of male full-time associates is 14% higher than that of female full-time associates (57% vs. 43%). For full-time professionals, there are three countries whose gender composition is unbalanced and leaning toward men – the Netherlands (71.9% vs. 28.1%), Germany (61.7% vs. 38.3%) and South Korea (57.5% vs. 42.5%). Somewhat surprisingly, the number rises to 10 for full-time associates. The gender gap in favor of men is especially large in the Netherlands (76% vs. 24%, or delta = 52%) and Japan (75% vs. 25%, or delta = 50%) and the smallest in Ireland (53% vs. 47%, or delta = 6%). To put this in perspective, the percentages of male associates are at least three times greater than the percentages of female associates in the Netherlands and Japan.

On the hand, the United Kingdom is the only country with an equal number of men and women working as full-time associates. What is more, the United States and two Visegrád Group countries (the Slovak Republic and Poland) and are the only three countries where a slightly larger proportion of female than male full-time associates was observed – 14%, 12% and 8% higher in percentage respectively. Note that in the sample of full-time professionals, these countries also boast a larger proportion of women than men – 9.4%, 17.8% and 5.8% higher in percentage respectively.

Figure 13. Gender Distribution of Full-time Associates (⩾ 25 years old) by Country (Figure 3b, corresponding to Figure 3)



Highest Level of Educational Attainment:

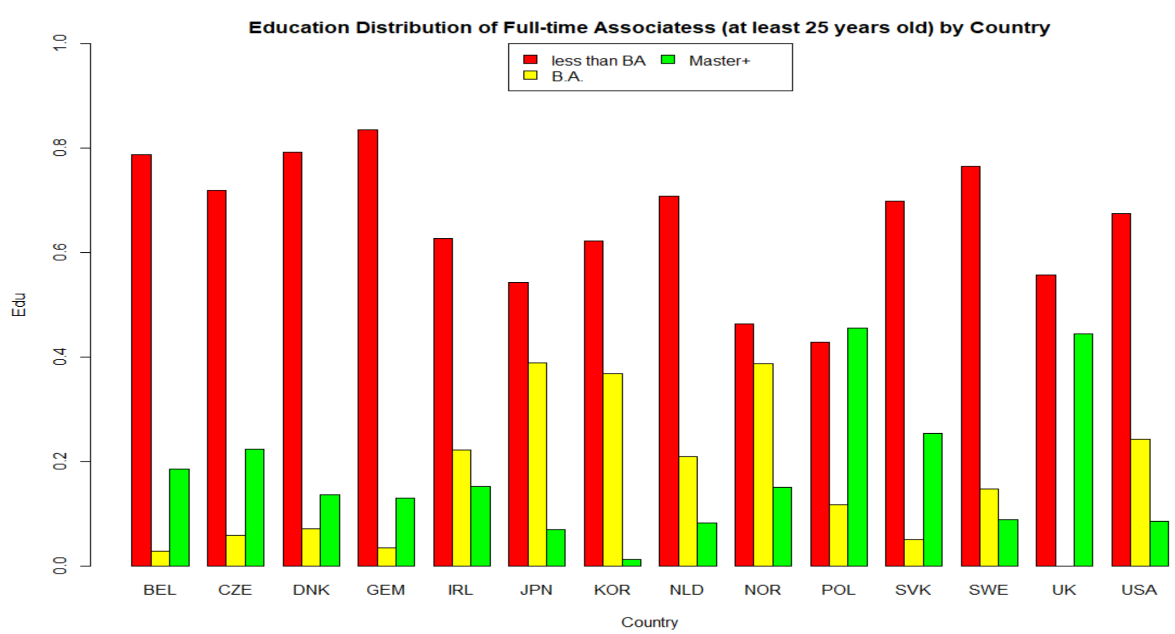
Compared to the sample of full-time professionals where only 30% have less than a Bachelor's degree, as high as 66% in the associated sample have less than a Bachelor's degree. Such differences are not unexpected considering the different levels of education required for different types of occupations. In terms of higher education, 16% of the sampled full-time associates reported having a Bachelor's degree only and 18% reported having a Master's degree or above, compared to 22.7% and 46.8% respectively in the sample of full-time professionals.

At the individual country level, Poland and Norway have the smallest share of full-time associates who have not completed a college education (42.8% and 46.3% respectively) while Germany, Denmark and Belgium have the largest share (83.5%, 79.2% and 78.7% respectively). Interestingly, a comparison between Figure 4 and Figure 14/Figure 4b reveals that although

Germany has the highest percentage of full-time associates without a Bachelor's degree, it also has the highest percentage of full-time professionals with at least a Bachelor's degree.

As is the case with the sample of full-time professionals, the two East Asian countries have the highest proportion of full-time associates who have a Bachelor's degree only (38.8% in Japan and 36.7% in South Korea), as well as the lowest proportion of full-time associates with a Master's degree or above (7% in Japan and 1.2% in South Korea). Similarly, the three Visegrád Group countries have the largest proportion of full-time associates who hold at least a Master's degree and a relatively small proportion of full-time associates with a Bachelor's degree only. However, the ratios of Master+ level and Bachelor level are not as extreme as those for full-time professionals; that is 5.2 to 1 in the Slovak Republic, 3.9 to 1 in Poland and 3.8 to 1 in the Czech Republic.

Figure 14. Education Distribution of Full-time Associates (⩾ 25 years old) by Country (Figure 4b, corresponding to Figure 4)



Highest Level of Parental Education:

In the associated sample, an average of 28% of full-time associates have at least one parent who has attained tertiary education, compared to 41.4% in the sample of full-time professionals. On the other hand, 26% of full-time associates reported having neither parent with a high school (upper secondary) education, not too far from 20% in the sample of full-time professionals. Figure 15 (or Figure 5b) shows that Germany and the United States are among countries boasting the highest percentage of tertiary-educated parents and the lowest percentage of parents without upper secondary education -- 37.8% vs. 5.3% and 36.3% vs. 14% respectively. However, the contrast is not as sharp as that among full-time professionals, i.e., 56.5% vs. 2.9% and 55.8% vs. 3.8% respectively.

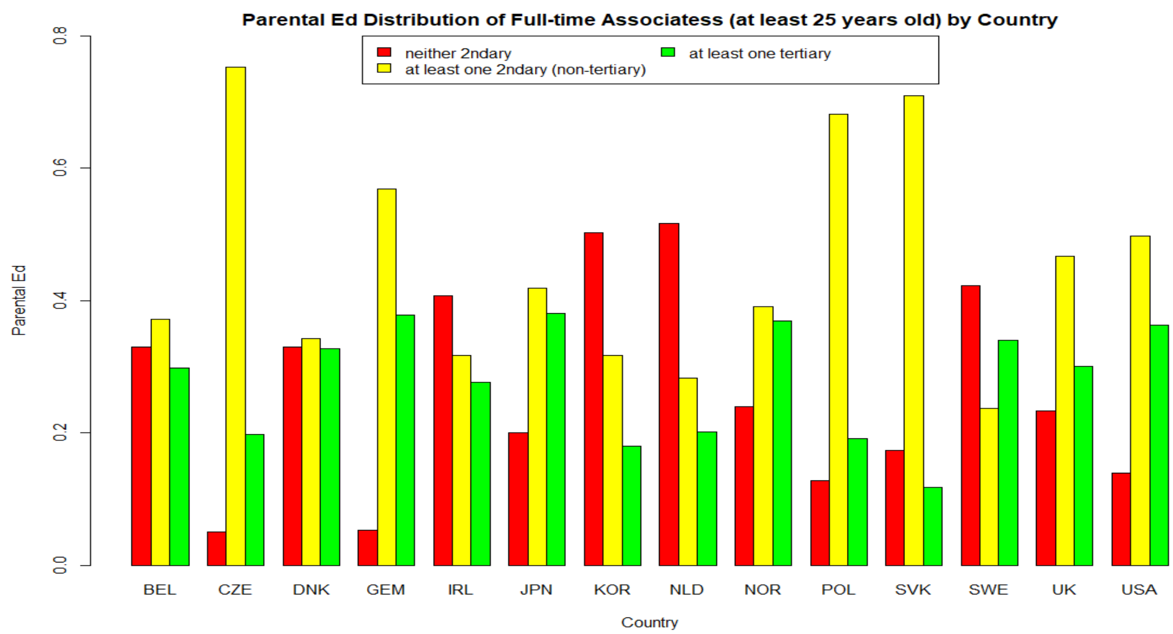
As is the case with full-time professionals, the American sample has the second-largest proportion of tertiary-educated parents but a smaller-than-average proportion of full-time associates with either a Bachelor’s or a Master’s degree. In other words, notwithstanding better-educated parents, the educational attainment of U.S. full-time associates in the sample does not compare favorably with that of their international competitors. This finding reinforces our concern that compared to their parents’ generations, the current U.S. workforce – both professionals and associates – may have lost the competitive edge in terms of the intergenerational mobility in education.

What is more, Japan and the three Scandinavian countries all have relatively high proportions of tertiary-educated parents. Similar to the sample of full-time professionals, these countries also reported much higher proportions of parents without upper secondary education than Germany and the United States. For instance, the proportion of non-tertiary-educated parents in these countries is four to eight times greater than the proportion of non-tertiary-educated parents in Germany (5.3%), ranging from 20% in Japan to 42.2% in Sweden.

On the other hand, the highest percentages of parents without upper secondary education were observed in the Netherlands (51.7%, compared with 39.9% for full-time professionals) and South Korea (50.2%, compared with 34.1% for full-time professionals). Again, note that although the two East Asian countries have roughly equal shares of full-time associates with at least a Bachelor's degree, the percentage of non-tertiary-educated parents is more than twice as large in South Kore as that in Japan (i.e., 50.2% vs. 20%). This finding further confirms the existence of a larger degree of intergenerational educational mobility in South Korea than in Japan.

Last but not least, the Visegrád Group countries reported the highest percentages of parents with upper secondary (non-tertiary) education and the lowest percentages of parents without upper secondary education – 75.3% vs. 5% in the Czech Republic, 68.1% vs. 12.7% in Poland and 70.9% vs. 17.3% in the Slovak Republic. Note that all these percentages are higher than their counterparts in the sample of full-time professionals. Or, put differently, the proportion of tertiary-educated parents is higher among full-time professionals than among full-time associates. More importantly, given the fact that these countries boast the largest proportion of Master+ level but the third-smallest proportion of Bachelor level full-time professionals (refer back to Figure 14/Figure 4b), an upward intergenerational mobility in education is evident within the Visegrád Group.

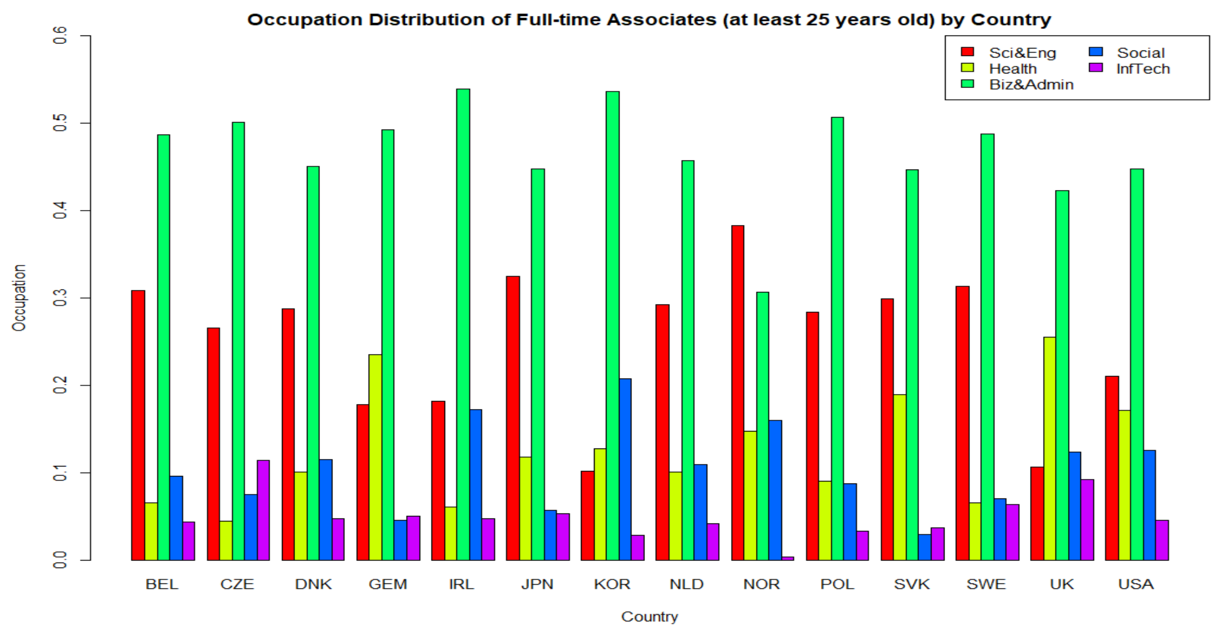
Figure 15. Parental Ed. Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 5b, corresponding to Figure 5)



Occupational Categories for Full-time Associates:

Different from the sample of full-time professionals where there are six occupational categories and Teaching professionals comprise about 30% of the sampled population before weighting, the associated sample has only five occupational categories (minus Teaching) and Business & Administration has the largest share of full-time associated professionals. On average across the 14 selected countries, almost half (46.1%) of the sampled full-time associates are working in the Business & Administration field. As shown in Figure 16 (or Figure 6b), Norway is the country with the lowest percentage of Business & Administration associates (30.7%) and the only country where Business & Administration is not the most dominant occupation (only second to Science & Engineering).

Figure 16. Occupation Distribution of Full-time Associates (≥ 25 years old) by Country (Figure 6b, corresponding to Figure 6)



Accounting for 25.5% of the sampled associates, Science & Engineering is the second-largest profession in most countries, except in Norway (38.3%), Germany (17.8%), the United Kingdom (10.6%) and South Korea (10.1%). Note that in the latter two countries, Science & Engineering associates make up the second-smallest proportion of the sampled full-time associates, only larger than the proportion of those working as Information & Communications Technicians.

In the sample of full-time professionals, unbalanced distributions of occupational categories were observed in a few countries where respondents from two of the six occupational categories (i.e., Teaching and Science & Engineering) represented almost two thirds of the sample. Such unbalanced distributions became more obvious when it comes to the associated sample. As mentioned before, Business & Administration alone takes up the lion's share in all but one country (i.e., Norway), ranging from 42% in the United Kingdom to 54% in Ireland and

Japan. Moreover, the combination of Business & Administration and Science & Engineering constitutes an absolute majority in all the 14 selected countries, ranging from 52.9% in the United Kingdom to 80.0% in Sweden.

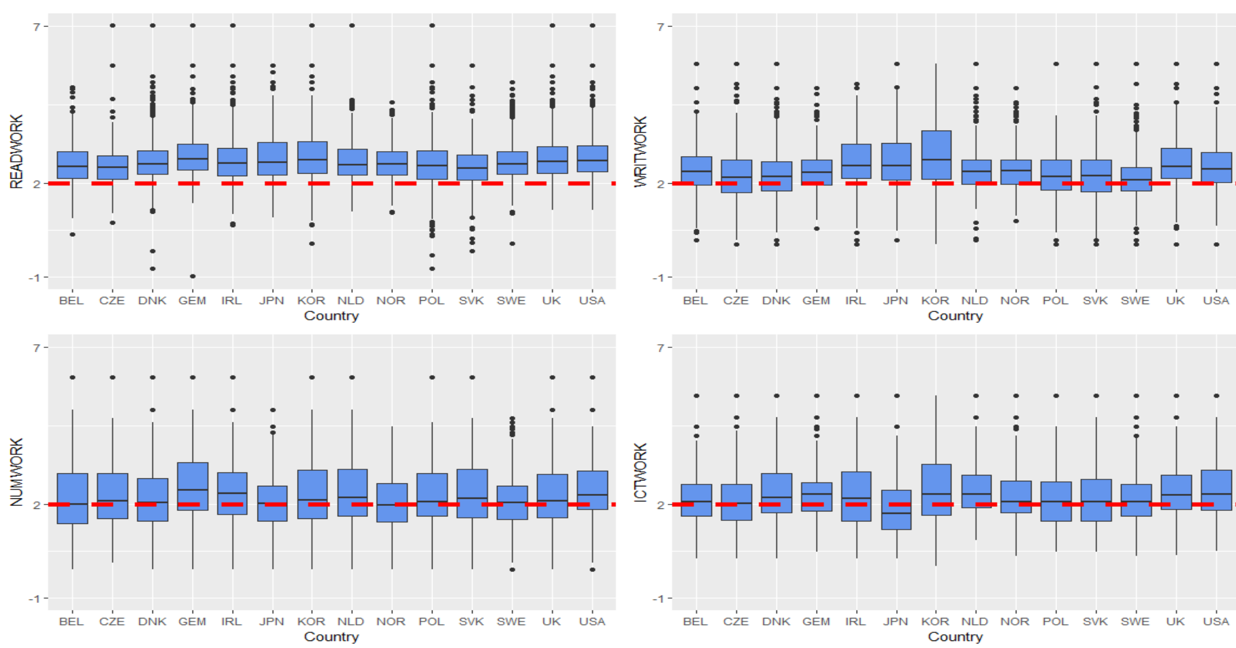
3.5.2. Skills Use at Work, at Home and Job-Related Activities of Sampled Full-time Associates in the 14 Selected Countries

Use of Key Information-Processing Skills at Work

Just as with full-time professionals, full-time associates in the sample reported Reading and Writing skills as the most frequently used key information-processing skills at work. However, full-time associates do not use Reading skills at work as frequently as full-time professionals. Among the 6,146 sampled full-time associates, the average use of Reading and Writing skills at work is 2.45 and 2.4 respectively, compared with 2.7 and 2.4 among full-time professionals. As demonstrated in Figure 17 (or Figure 7b), the average use of both Literacy skills is well above the grand mean in all the 14 selected countries. Full-time associates in Norway and Japan reported using Reading skills most frequently at work (>.55 standard deviations above 2), followed by those in Germany and the United Kingdom (2.55). Note that these frequencies are at least .3 standard deviations less than what was reported by full-time professionals in South Korea and Germany. The least frequent use of Reading skills at work is again observed in the three Visegrád Group countries (< .3 standard deviations above 2), which is also .3 standard deviations less than what was reported by full-time professionals in these same countries. On the other hand, full-time associates reported using Writing skills at work in a similar manner as full-time professionals, in terms of the intensity of use and the distribution across countries. Writing skills are used most frequently in the United Kingdom (> .65 standard deviations above 2) and least in two of the Scandinavian countries (2.07 and 2.26 in Sweden and

Denmark respectively). The three Visegrád Group countries also reported relatively low frequencies of Writing skills use at work -- approximately 2.35, compared with 2.28 in the sample of full-time professionals. To sum up, among the sampled full-time associates, both Literacy skills are used most intensively at work in the United Kingdom (vs. South Korea for full-time professionals) and least in the three Visegrád Group countries.

Figure 17. Distribution of the Use of Key Information-Processing Skills at Work (Figure 7b, corresponding to Figure 7)



As far as the use of Numeracy skills at work is concerned, both full-time associates and full-time professionals reported average use equals the grand mean. As is the case with the sample of full-time professionals, the average use of Numeracy skills at work is greater than the grand mean in all the 14 selected countries. However, at the individual country level, the distribution of the use of Numeracy skills at work varies considerably between the two samples. First, full-time associates in the Czech Republic reported using Numeracy skills most frequently at work (2.59), followed by those in the United States and the United Kingdom (2.38 and 2.36

respectively). Recall that in the sample of full-time professionals (refer back to Figure 7), the Czech Republic was ranked among the bottom six countries due to its relatively low frequency of Numeracy skills use at work (2.24). Second, in the associated sample, Germany is the country with the third-lowest average use of Numeracy skills at work (2.33), only higher than South Korea and Norway (2.20 and 2.19 respectively). In contrast, full-time professionals in Germany reported the highest frequency of using Numeracy skills at work (2.53).

It goes without saying that the two samples share a number of similarities. For instance, the United States is the country with the second-highest average use of Numeracy skills at work in both samples, while Norway remains the country where the lowest frequency is reported. You may also notice that, within the associated sample, the United Kingdom is the only country consistently ranked at the top for its intensive use of both Literacy and Numeracy skills in the workplace.

When it comes to the use of ICT skills at work, the sampled full-time associates reported a lower frequency than full-time professionals (2.0 vs. 2.3) across countries. Similar to the sample of full-time professionals, the average use is above the grand mean in all but one country (i.e., Japan). Note that in both samples, Japan is not only the country where ICT skills are used the least at work, but also the only one whose average use is slightly below the grand mean (1.99 for full-time associates and 1.87 for full-time professionals). For individual countries, ICT skills are used most frequently at work in Ireland, Denmark and the United Kingdom (>.35 standard deviations above 2) and least in Japan, followed by Germany, the Slovak Republic and Poland (< .15 standard deviations above 2).

To sum up, for full-time associates, a few findings have emerged from comparing the use of key information-processing skills at work across the 14 selected countries:

(1) The United Kingdom was consistently ranked at the top for its relatively high frequency of using of all the key information-processing skills in the workplace. Germany is the country ranked high in the use of Reading skills at work but low in Numeracy and ICT skills.

(2) South Korea is the country with the second-lowest average use of Numeracy skills at work while Japan is the country with the lowest average use of ICT skills at work.

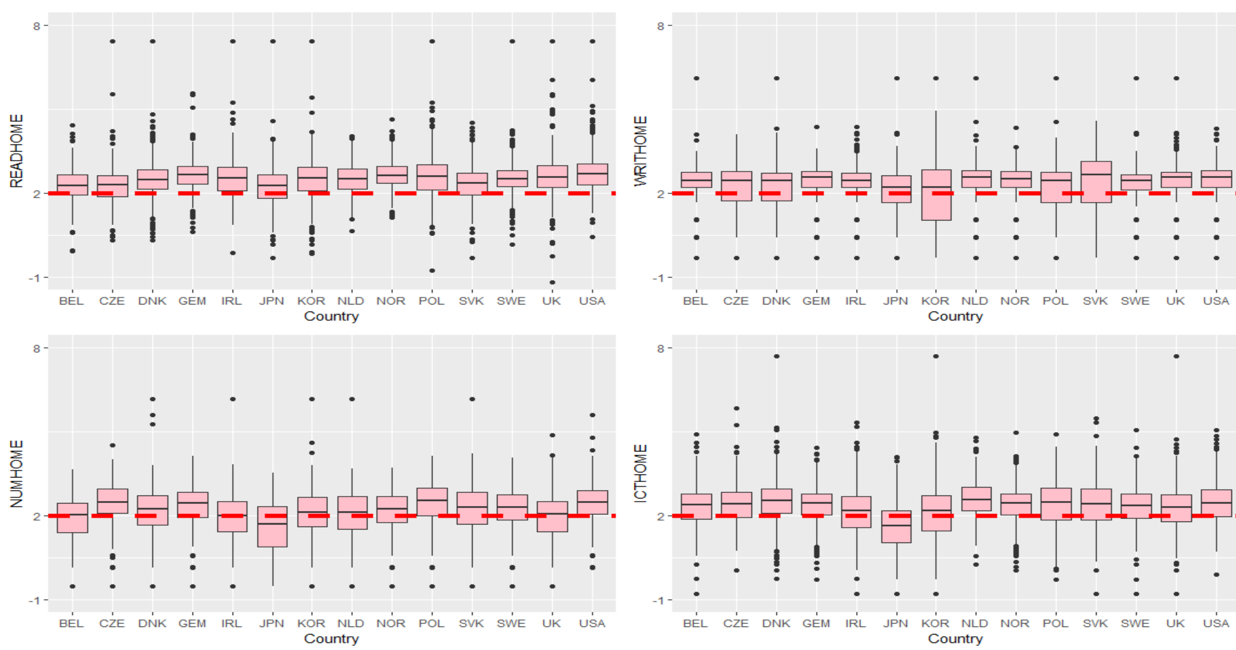
(3) Full-time associates in the three Visegrád Group countries did not report using Literacy and ICT skills at work as intensively as their international peers. The same observation applies to the sampled full-time professionals.

(4) Just as with the sample of full-time professionals, the use of Numeracy skills at work differed from that of Literacy and ICT skills among full-time associates.

Use of Key Information-Processing Skills at Home

As for the use of key information-processing skills at home, the sampled full-time associates also reported using Reading and Writing skills most frequently -- but in a lower frequency than full-time professionals. Across the 14 selected countries, the average use of Reading and Writing skills at home is 2.3 and 2.1 among full-time associates, compared with 2.5 and 2.3 among full-time professionals. It is worth noting that based on the frequency of both Literacy skills use at home, top-ranked countries in the sample of full-time professionals also ranked at the top with the associated sample. As shown in Figure 18 (or Figure 8b), The United States, Norway, the United Kingdom and Germany are the top 4 countries with the highest frequencies of Reading skills use at home (>.45 standard deviations above 2 for full-time associates vs. > .60 standard deviations above 2 for full-time professionals). While the Netherlands, the United Kingdom, the Slovak Republic, the United States and Germany are the top 5 countries with the highest frequencies of Writing skills use at home (>.18 standard deviations above 2 for full-time associates vs. > .36 standard deviations above 2 for full-time professionals). By contrast, Japan has the lowest level of Reading skills use at home (2.06, compared with 2.24 for full-time professionals) while South Korea has the lowest level of Writing skills use at home (1.83, compared with 1.98 for full-time professionals). Note that South Korea is the only country in the sample of full-time professionals with a below-the-grand-mean use of Writing skills at home; however, the number of such countries increased to four in the associated sample, including South Korea, Poland, Sweden and Japan (1.83, 1.87, 1.95 and 1.96 respectively).

Figure 18. Distribution of the Use of Key Information-Processing Skills at Home (Figure 8b, corresponding to Figure 8)



The Czech Republic is exceptional in the associated sample in that it has the second-lowest frequency of Reading skills use at home (2.13, only higher than 2.06 in Japan) but the highest frequency of Numeracy and ICT skills use at home (2.5 and 2.4 respectively). Following the Czech Republic, the other two Visegrád Group countries and the United States also reported relatively high levels of Numeracy skills use at home (>.15 standard deviations above 2). Japan remains the country where the lowest level of Numeracy skills use at home is observed (1.58), followed by the Netherlands, Ireland, Belgium and the United Kingdom (1.71, 1.74, 1.82 and 1.85 respectively). Also note that there are as many as five countries in the associated sample reported a below-the-grand-mean use of Numeracy skills at home, compared with four such countries (minus the Netherlands) in the sample of full-time professionals. In both samples, South Korea is the country that was ranked just above the bottom five due to its relatively low average use of Numeracy skills at home (2.0 for full-time associates vs. 2.06 for full-time professionals).

In addition to the Czech Republic, ICT skills are used most frequently at home in the Netherlands, Denmark and Norway (>.25 standard deviations above 2 for full-time associates vs. > .50 standard deviations for full-time professionals). On the other hand, the two East Asian countries reported the lowest average use of ICT skills at home (1.52 in Japan and 1.81 in South Korea). In short, such distributions are very similar to what was found in the sample of full-time professionals, notwithstanding the fact that ICT skills are used less intensively at home among full-time associates.

In summary, comparing the use of key information-processing skills at home between the two samples has yielded the following findings:

(1) Different from the sample of full-time professionals where three countries (i.e., Germany, the Netherlands and the United States) were ranked consistently near the top in the average use of all key information-processing skills at home, no such pattern was found in the associated sample. However, it might be worth mentioning that American full-time associates reported relatively

high average use of both Literacy and Numeracy skill at home but relatively low average use of ICT skills at home (2.21).

(2) Across the four skill domains, full-time associates in the two East Asian countries reportedly used all the key information-processing skills the least frequently at home, which is consistent with the observation from the sample of full-time professionals.

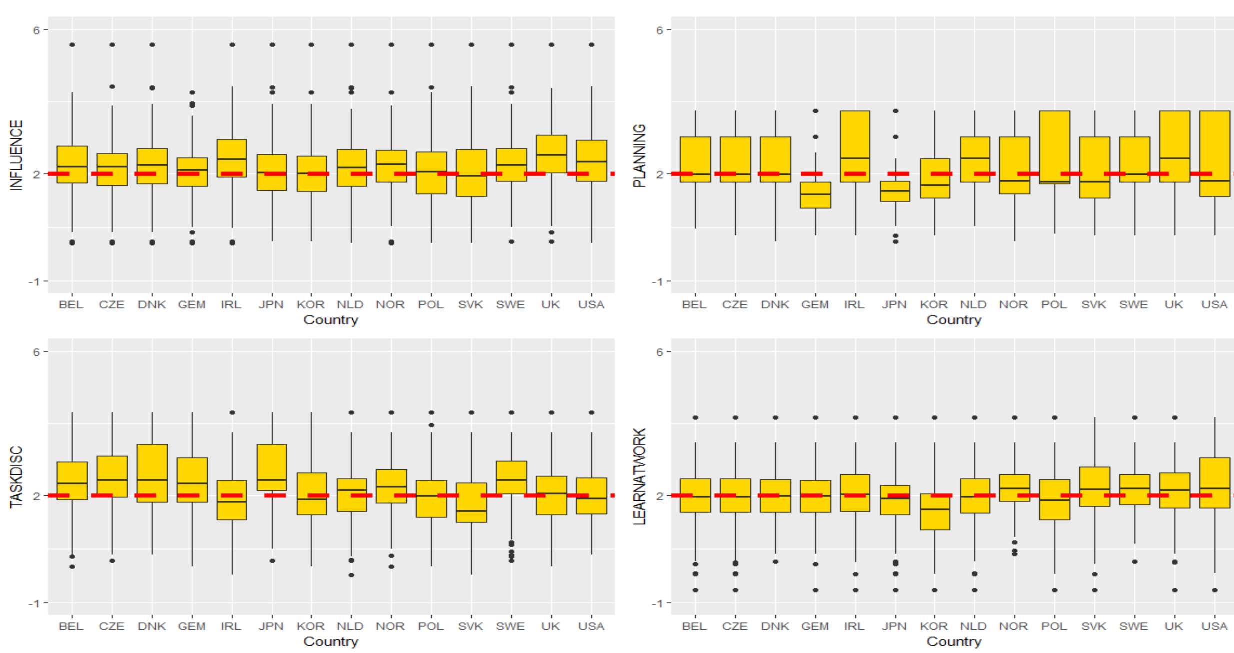
(3) Full-time workers – both professionals and associates – in the three Visegrád Group countries reported using Numeracy and ICT skills at home more intensively than their international peers. This is in sharp contrast to the previous finding that full-time workers in these countries have a less intensive use of key information-processing at work.

Use of Generic Workplace Skills/Job-related Activities

As is the case with full-time professionals, the use of generic workplace skills and the country rankings among full-time associates vary substantially by skills domain. Across the 14 selected countries, influencing skills remain to be seen as the most frequently used generic workplace skills; however, the average use among the sampled full-time associates is .2 standard deviations lower than that of full-time professionals (2.2 vs. 2.4). The United Kingdom, Ireland and the United States are still the countries where influencing skills are used most frequently at work (>.35 standard deviations above 2 for full-time associates vs. > .70 standard deviations above 2 for full-time professionals). While Japan, along with South Korea and the Slovak Republic, are countries where influencing skills are use least frequently (approximately 2.0).

The use of Planning skills in the workplace continues to vary widely within and across the 14 selected countries. The sampled full-time associates reported 2.2 as the average use of Planning skills at work, which is .1 standard deviations less than what was reported by full-time professionals. The highest frequency of Planning skills use at work was observed in the United Kingdom and Ireland (>.45 standard deviations above 2 for full-time associates vs. > .70 standard deviations above 2 for full-time professionals) while the lowest is in the Germany, Japan and South Korea (1.62, 1.72 and 1.90 respectively). As evidenced in Figure 19 (or Figure 9b), the distribution of Planning skills use in the workplace displays great variability within and across the 14 selected countries. Similar to Figure 9, the United States, Ireland and the United Kingdom are countries with the largest interquartile range (IQR) while Germany and Japan are countries with the smallest IQR. What is more, Germany and Japan are the two countries whose 75th percentile falls below the grand mean completely. In fact, there is no major difference between the two samples regarding the use of Planning skills in the workplace.

Figure 19. Distribution of the Use of Workplace Skills/Job-related Activities (Figure 9b, corresponding to Figure 9)



Across countries, there is no difference between the sampled full-time associates and full-time professionals in the average use of task discretion in the workplace (i.e., 2.3 for both samples). At the individual country level, the same four countries reported using Task Discretion skills most frequently at work. Note that it is the first time that countries in the associated sample have a higher average use of skills at work than their counterparts in the sample of full-time professionals. Specifically, full-time associates in Japan, Sweden, Denmark and Germany reportedly used Task Discretion skills more intensively than full-time professionals in these countries -- 2.67 vs. 2.50, 2.59 vs. 2.48, 2.51 vs. 2.49, and 2.48 vs. 2.47 respectively. On the other hand, the Slovak Republic and Ireland remain the countries with the lowest average use of Task Discretion skills at work (1.84 and 1.92, compared with 1.99 and 1.89 in the sample of full-time professionals). As is the case with full-time professionals, full-time associates in Ireland reported the most intensive use of Influencing and Planning skills, but the least use of Task Discretion in the workplace.

As for learning at work, the United States emerges as the country with the highest frequency of Learning at Work (2.30), followed by the Slovak Republic (2.25) and the three Scandinavian countries (ranging from 2.24 in Norway to 2.11 in Denmark). On the other hand, Japan remains the country with the third-lowest average use of Task Discretion skills at work, followed by the Czech Republic and South Korea (1.86 and 1.59 respectively).

Skills Used Jointly in the Workplace

Regarding the clustering of skills used at work, the associated sample yields results resembling that from the sample of full-time professionals. Influencing and Planning skills continue to have the strongest correlation in all the 14 selected countries. Germany and the United Kingdom remain the countries where the lowest and the highest correlations were observed (.46 and .61 respectively). The second strongest correlation was still found between Reading and ICT skills at work, ranging from .32 in Norway to .51 in the Netherlands. It is interesting to notice that the Reading-ICT correlation in Norway is the highest with the sample of full-time professionals but the lowest with the associated sample. Lastly, a careful examination of the correlation matrices in Appendix F confirmed that for both samples, influencing skills are more closely related with Reading and Writing skills than any other workplace skills. In other words, the clustering of skills used at work among the sampled full-time associates follows the same pattern as that among full-time professionals, regardless of the different occupational compositions in the two samples.

3.5.3 The Different Formats of AET Participation for Sampled Full-time Associates in the 14 Selected Countries

As you may recall in Section 3.4.3, among the 14 selected countries in the entire PIAAC sample, only five countries reported an AET participation rate greater than 60%, aka, Group 1 (> 60%) in Table 11. While in the sample of full-time professionals only (at least 25 years old), the AET participation rate exceeded 64% in all the 14 selected countries. Therefore, we heightened the grouping standards and reclassified all these countries into four new groups. Please refer to Table 11 for the new grouping of countries based on the heightened standards.

In this section, we applied the heightened standards to the associated sample and compared the grouping of countries between the two samples (see the last two columns of Table 11). The first thing to be noticed is that no countries in the associated sample have a participation rate higher than 85%, compared with four countries in the sample of full-time professionals. Notwithstanding the absence of the New Group 1 (>85%), the rankings of countries is very similar to that based on the sample of full-time professionals, with the exception of two countries: United States and Denmark. For full-time professionals aged 25-65, both countries enjoyed the highest AET participation rate (89% and 88% respectively) among the 14 selected countries. When it comes to full-time associates of the same age range, the participation rate decreased to 76% and 77% respectively, which are only marginally higher than the average of the associated sample (i.e. 73%). Expressed in a different way, the U.S. and Denmark rankings dropped two notches (i.e., from Group 1 to Group 3) and only took the 8th and the 5th spots respectively within the associated sample. On the other hand, the Netherlands and Sweden continue to boast the highest AET participation rate with full-time associates (81%, compared to 87% with full-time professionals) while the Slovak Republic is still the country where the lowest participate rate for AET programs is observed (58.3%, compared to 64% in the sample of full-time professionals).

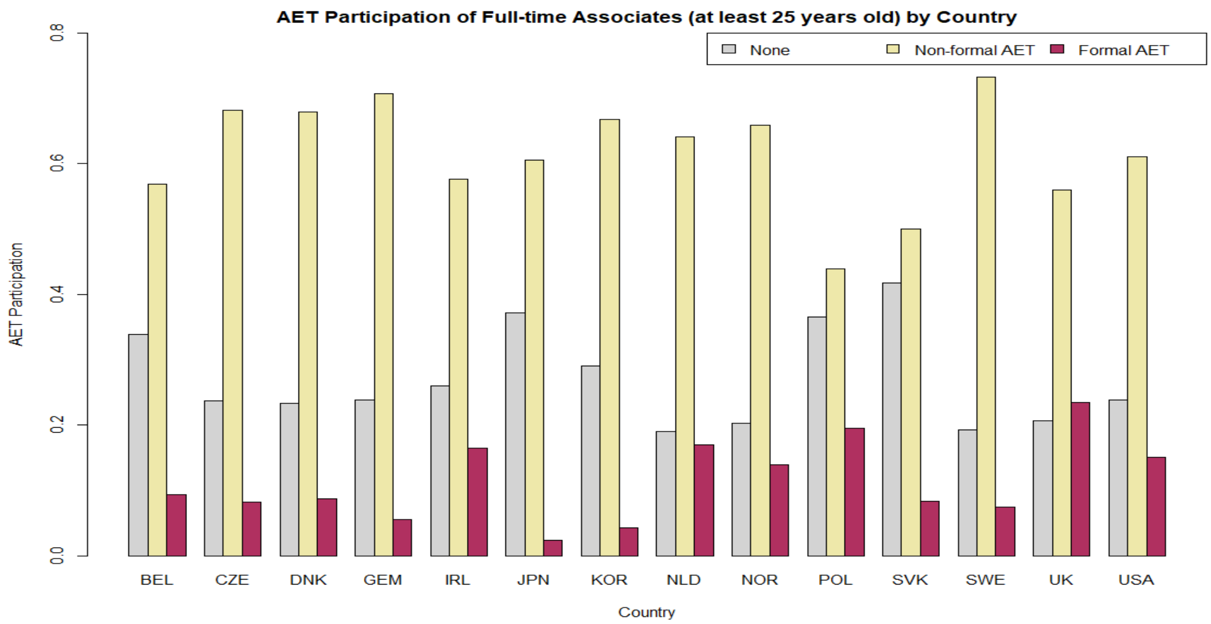
In terms of the different formats of AET participation, Non-formal AET programs are still the most popular among the sampled full-time associates. The average Non-formal AET participation rate is as high as 62% in the associated sample, which is 5% lower than the sample of full-time professionals. The Non-formal AET participation rate among full-time associates is the highest in Sweden (73.3%), followed by Germany, the Czech Republic and Denmark (70.7%, 68.2% and 67.9%). On the low end, Poland has replaced the Slovak Republic as the country with the lowest participation rate for Non-formal AET programs (43.8%).

In the associated sample, the proportion of non-participants is more than double the size of those who participated in formal AET programs during the year prior to the survey – 26.8% vs. 11.5%, compared with 17.8% vs. 15.5% in the sample of full-time professionals. From another perspective, the average non-participation rate in the associated sample is 9% higher than that in the sample of full-time professionals, while the average formal AET participation rate is 4% lower than that in the sample of full-time professionals.

As illustrated in Figure 20 (or Figure 10b), the Netherland and Sweden are countries with the lowest percentages of non-participants (approximately 19%); while the Slovak Republic, Japan, Poland and Belgium are countries where the highest percentages of non-participants were observed (41.7%, 37.1%, 36.6% and 33.9% respectively). Three finding have emerged from examining the country rankings of non-participation rate for both samples. First, while the United States boasts the lowest non-participation rate among full-time professionals, the non-participation rate among full-time associates is not as extreme. In the associated sample, the United States only ranked in the middle of the pack with a non-participation rate higher than 6 countries including the Czech Republic. Second, for both full-time professionals and full-time associates, the Czech Republic has the lowest non-participation rate among the three Visegrád Group countries. Third, unlike Japan whose non-participation rate is the second-highest in both samples, the non-participation rate in South Korea is close to the sample average for both types of full-time workers.

On the other hand, as high as 23.4% of full-time associates in the United States reported having participated in formal AET programs in the previous 12 months, followed by 19.6% in Poland. The lowest participation rate for formal AET programs is observed in Japan (2.3%) and South Korea (4.2%). Again, three findings have emerged from comparing the country rankings for formal AET participation rate and non-participation rate. First, notwithstanding a mediocre participation rate for AET programs in general, full-time associates in the United States boast the highest participation rate for formal AET programs in particular. Second, both types of full-time workers in Poland reported the highest formal AET participation rate among the 14 selected countries. Third, the two East Asian countries are similar when it comes to the formal AET participation rate among full-time associates, that is, both countries have the lowest formal AET participation rate among the 14 selected countries.

Figure 20. Distribution of AET Participation among Full-time Associates (≥ 25 years old) by Country (Figure 10b, corresponding to Figure 10)



Chapter 4. Results

4.1 Research Question One

Research question one concerns describing full-time professionals' levels of PS-TRE proficiency scores and comparing distributions of proficiency scores across the 14 selected countries. Within each country, the distributions of full-time professionals' PS-TRE proficiency scores are examined by gender and age. What is more, research question one explores the relationships between full-time professionals' proficiency scores in PS-TRE and their proficiency scores in Literacy and Numeracy. Lastly, research question one compares the distributions of full-time professionals' PS-TRE proficiency scores with those of full-time associated professionals. Specifically, the following four sub-questions are answered for research question one.

4.1.1 RQ1a

RQ1a asks about the distributions of full-time professionals' PS-TRE proficiency scores across the 14 selected countries that participated in the first round of PIAAC. A series of descriptive statistics were obtained for each country. As shown in Figure 21, the mean PS-TRE proficiency score (including the 95% confidence interval (CI) of the mean), the interquartile range (IQR) as well as the 5^{th} and 95^{th} percentiles were calculated for each country based on the 10 plausible values provided by the PIAAC Consortium (von Davier, Gonzalez & Mislevy, 2009). Additionally, to account for the complex sample design applied in PIAAC, replicate weights were assigned to each individual.

Figure 21. The Distributions of Full-time Professionals PS-TRE Proficiency Scores by Country

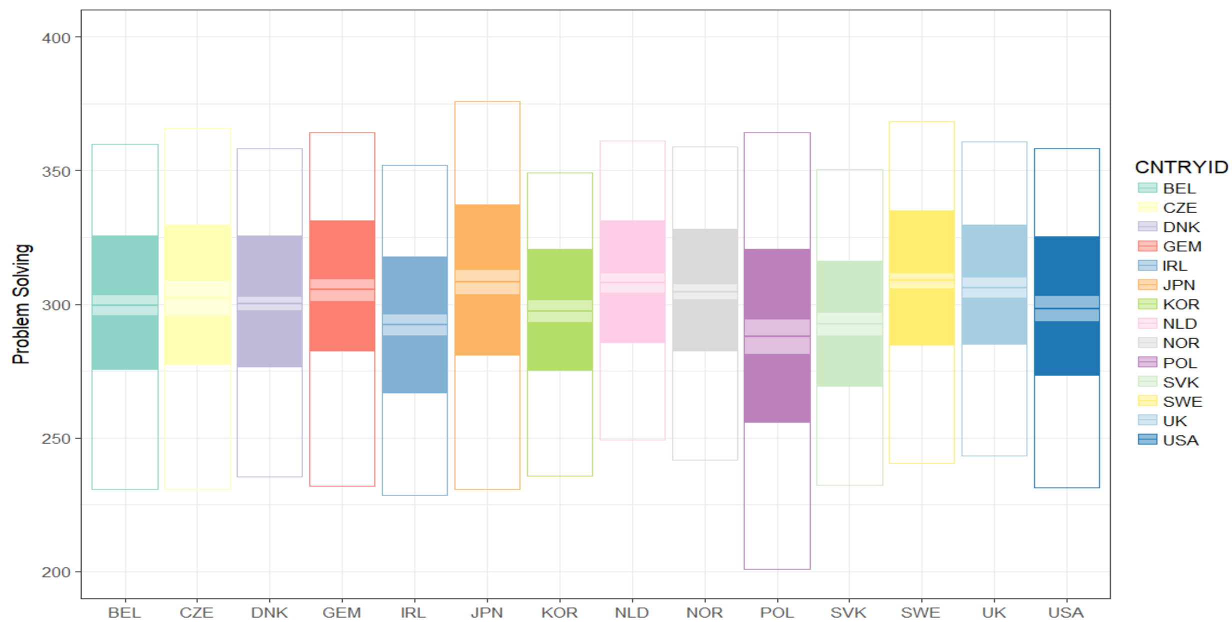


Table 12. The Rankings of Countries by the Mean PS-TRE Proficiency Score for Full-time Professionals

CNTRYID	# of Cases	Mean Score	s.e.	80th percentile	85th percentile	90th percentile
Sweden	*681*	309	1.5	**340**	**347**	**355**
The Netherlands	*462*	308	2.0	337	**343**	**350**
Japan	*369*	308	2.5	**343**	**351**	**361**
The United Kingdom	*743*	306	2.1	335	**342**	**348**
Norway	*659*	305	1.5	333	339	**347**
Germany	*512*	305	2.2	338	**345**	**354**
The Czech Republic	*392*	302	3.4	336	**343**	**353**
Denmark	*1,190*	300	1.4	331	338	**347**
Belgium	*509*	300	3.0	332	339	**347**
The United States	*560*	298	2.5	331	338	**347**
South Korea	*479*	297	2.2	325	332	338
The Slovak Republic	*367*	293	2.3	322	329	337
Ireland	*619*	292	2.1	323	331	340
Poland	*389*	288	3.4	329	337	**347**

While Figure 21 presents graphic comparisons of the results across the 14 selected countries, Table 12 ranks these countries by the mean scores (in descending order) and then the standard errors of the mean for ties (in ascending order). As it turns out, Sweden has the highest mean PS-TRE score and the second-lowest standard error of the mean (mean = 309, s.e. = 1.5), followed by the Netherlands and Japan (mean = 308, s.e. = 2.0 and 2.5 respectively); while Poland has the lowest mean PS-TRE score and the highest standard error of the mean (mean = 288, s.e. = 3.4). Because proficiency scores have been split into different "proficiency levels" in the official report (OECD, 2013c), it is important to know that on average, full-time professionals (at least 25 years old) in all the 14 selected countries performed at Level 2/ "moderate" (scores from 291 to 340) in PS-TRE. That is to say, on average, the sampled respondents "can complete problems that have explicit criteria for success, a small number of applications, and several steps and operators" (OECD, 2013c). However, none of the 14 selected countries boast a mean PS-TRE score at the highest level (Level 3 with scores higher than 340) where adults "can complete tasks involving multiple applications, a large number of steps, impasses, and the discovery and use of ad hoc commands in a novel environment" (OECD, 2013c).

In terms of the percentage of Level 3/ "strong" performers in PS-TRE, Japan and Sweden are the two countries with the highest percentages of full-time professionals (at least 25 years old) scoring higher than 340 – 24% and 21% respectively. While the Slovak Republic and South Korea are the two countries with the lowest percentages of Level 3/ "strong" performers – 8% and 10% respectively.

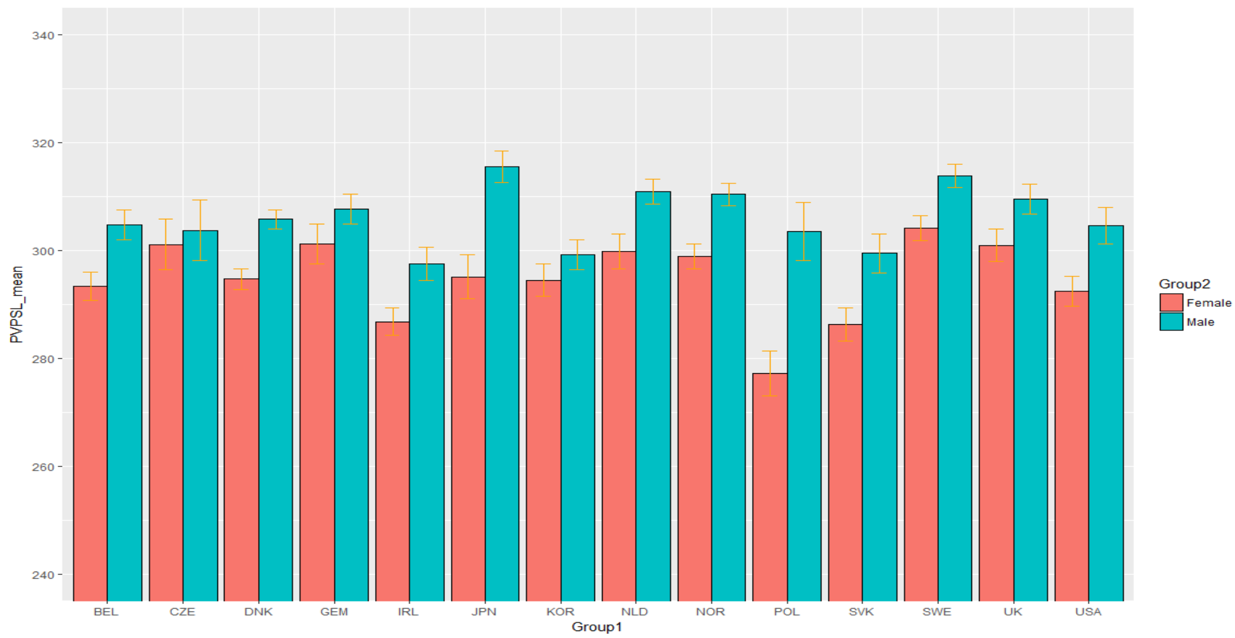
If we look at the 80th percentile of the score distribution for each country, Sweden and Japan are again the two countries whose full-time professionals performed at Level 3. That is to say, in Sweden and Japan, full-time professionals at or above the 80th percentile of the score distribution "can complete tasks involving multiple applications, a large number of steps, impasses, and the discovery and use of ad hoc commands in a novel environment". They can also "establish a plan to arrive at a solution and monitor its implementation as they deal with unexpected outcomes and impasses" (OECD, 2013c). Moving up to the 85th percentile for each country, the number of countries performing at Level 3 increases to six. At the 90th percentile, almost all countries have achieved the highest level of performance in PS-TRE, with the exception of Ireland, South Korea and the Slovak Republic.

Further examinations of Table 12 suggest that the country rankings may differ at different percentiles of the score distributions. For example, Japan, rather than Sweden, Japan stands out as the country with the highest PS-TRE proficiency scores across the upper percentiles. Poland also outperforms the aforementioned three countries -- Ireland, South Korea and the Slovak Republic -- in the upper percentiles despite having the lowest mean PS-TRE score.

4.1.2 RQ1b

What is more, it is necessary to check the distributions of PS-TRE proficiency scores according to full-time professionals' demographic background factors that are comparable across countries. In RQ1b, we focus on the mean PS-TRE score for each unique combination of Gender*Age groups for each country and identify which particular combinations of Gender*Age groups have the highest PS-TRE proficiency scores. Figure 22 illustrates the variations of mean PS-TRE proficiency scores related to gender within and across the 14 selected countries.

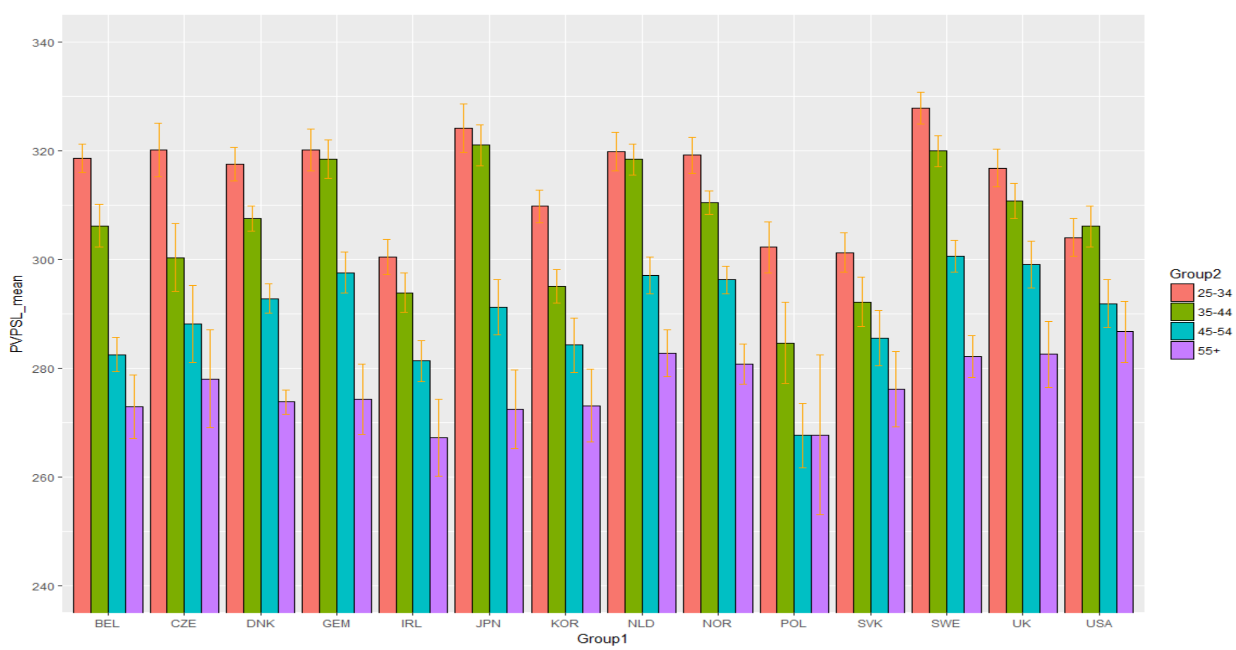
Figure 22. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Gender and Country



For male full-time professionals, the highest mean PS-TRE scores were obtained by Japan (mean = 316, s.e. = 2.9) and Sweden (mean = 314, s.e. = 2.2) while the lowest was observed in Ireland (mean = 298, s.e. = 3.0), South Korea (mean = 299, s.e. = 2.7) and the Slovak Republic (mean = 300, s.e. = 3.6). For female full-time professionals, the highest mean PS-TRE score was obtained by Sweden (mean = 304, s.e. = 2.3) while the lowest was observed in Poland (mean = 277, s.e. = 4.2). In other words, regardless of gender, full-time professionals in Sweden outperformed their international peers in PS-TRE as measured by PIAAC. But Japanese women did not stand out as much as Japanese men in PS-TRE. In fact, Japanese women scored as low as South Korean women on the PS-TRE scale, but Japanese men outscored South Korean men by as much as 17 points (s.e. = 2.9 in Japan and s.e. = 2.7 in South Korea), the largest between-country difference observed among male full-time professionals.

To delve deeper into gender differences in PS-TRE proficiency scores across countries, Figure 22 shows that men scored higher than women in all 14 countries. The gap is the greatest in Poland (delta = 27, s.e. = 3.4) and Japan (delta = 21, s.e. = 2.5) and the smallest in the Czech Republic (delta = 3, s.e. = 3.4) and South Korea (delta = 4, s.e. = 2.2). Note that Poland and Japan are also countries with the lowest and the highest mean PS-TRE scores respectively. South Korea is one of the three countries with the lowest PS-TRE proficiency scores across the upper percentiles (see Table 12). Therefore, it appears that the two East Asian countries are quite different not only in the level of PS-TRE proficiency among full-time professionals, but also in the extent of the gender gap. The same observation applies for the Visegrád Group where the Czech Republic tends to have higher PS-TRE scores but a smaller gender gap than the other two countries.

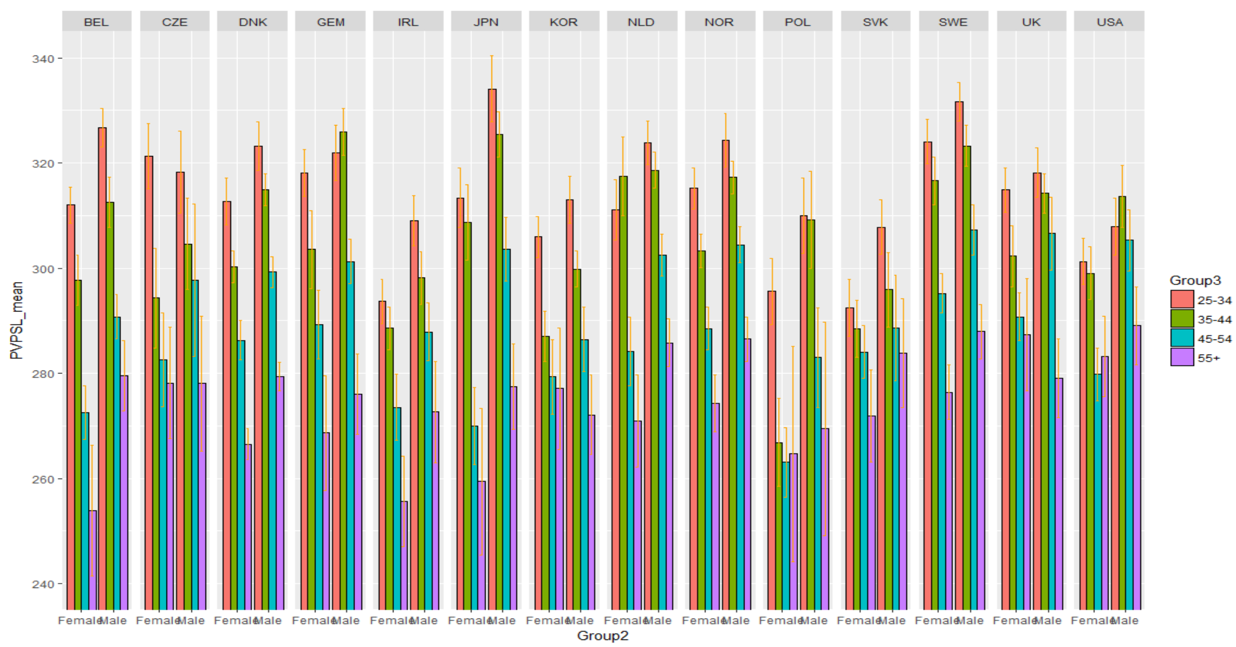
Figure 23. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Age and Country



On the other hand, Figure 23 illustrates the variations of mean PS-TRE proficiency scores related to age within and across the 14 selected countries. For the oldest cohort (i.e., 55-65 year olds) in the sample, the highest mean PS-TRE proficiency score was obtained by the United States (mean = 287, s.e. = 5.6) while the lowest is again observed in Ireland (mean = 267, s.e. = 7.0) and Poland (mean = 268, s.e. = 14.7). Note that scores for the oldest cohort in the two East Asian countries are also relatively low (mean = 272, s.e. = 7.3 in Japan, and mean = 273, s.e. = 6.7 in South Korea). For the youngest cohort (i.e., 25-34 year olds), the highest mean PS-TRE score was obtained by Sweden (mean = 328, s.e. = 3.0) and Japan (mean = 324, s.e. = 4.0) while the lowest was observed in Ireland (mean = 300, s.e. = 3.2), closely followed by the two Visegrád Group countries (mean equals 301 in the Slovak Republic and 302 in Poland). Perhaps most shocking among these findings is that scores for the youngest cohort in the U.S. are disappointingly low (mean = 304, s.e. = 3.5) -- only marginally higher than the aforementioned three countries ranked last, but considerably lower than the other 10 countries in the sample.

With regard to age-related differences in PS-TRE proficiency scores, Figure 23 demonstrates that mean scores decline from younger to older cohorts in all 14 countries. As pointed out in the OECD document (2013c), the reason for this is twofold: (1) the effects of biological ageing and (2) cohort effects which refer to “differences in the amount and quality of the opportunities that individuals have had to develop and maintain proficiency (particularly, but not exclusively, through formal education and training) over their lifetimes”. The gap between the youngest and the oldest cohort is the greatest in Japan (delta = 52, s.e. = 2.5) and the smallest in the United States (delta = 17, s.e. = 2.5) and the Slovak Republic (delta = 25, s.e. = 2.3). For the United States, such results are cause for concern. For one thing, the fact that the PS-TRE proficiency of young professionals is little different from their parents’ generations signals poor intergenerational progress in the U.S., especially in comparison to other countries. For another, the relatively low PS-TRE scores for the youngest cohort suggests that U.S. young professionals have lost the advantage to their international competitors.

Figure 24. Distribution of Full-time Professionals PS-TRE Proficiency Scores by Gender, Age and Country



Lastly, when the sample was split by both gender and age cohort simultaneously, the age-related declining trend persists in both genders across all 14 countries with few exceptions. Among these gender*age subgroups, 25-34-year-old men in Japan, Sweden and Belgium scored the highest on the PS-TRE scale (334, 332 and 327 respectively) while 55-65-year old women in Belgium, Ireland and Japan scored the lowest (254, 256 and 259 respectively). The greatest gender*age gaps were observed between the youngest male and the oldest female subgroups in Japan (delta = 75, s.e. = 2.5), followed Belgium (delta = 73, s.e. = 3.0). The smallest gender*age gaps were observed between the same two subgroups in the United States (delta = 25, s.e. = 2.5), followed by the United Kingdom (delta = 31, s.e. = 2.1). For the most part, the results align nicely with what have was found when the sample was split by gender or age separately.

The United States and Poland are the two countries where a few exceptions were noticed. In the U.S., the mean PS-TRE score for the 35-44-year-old men is 6 points higher than that of the youngest cohort (delta = 314 – 308, s.e. = 2.5), while the mean score for the oldest female group is 3 points higher than that of their 45-54-year old peers (delta = 283 – 280, s.e. = 2.5). In Poland, 35-44-year-old men scored almost the same as the youngest cohort (delta = 310 – 309, s.e. = 3.4), while the oldest female group scored 2-point higher than their 45-54-year old peers (delta = 265 – 263, s.e. = 3.4). However, note that many of these estimates come with large standard errors due to small sample sizes. Therefore, the interpretability of some results at the gender*age level is open to question.

4.1.3 RQ1c

RQ1c examines the correlations between full-time professionals' PS-TRE proficiency scores and their proficiency scores in Literacy and Numeracy. On average across the 14 selected countries, the correlation between proficiency scores in PS-TRE and Literacy at the individual level is .79, and the correlation between proficiency scores in PS-TRE and Numeracy at the individual level is .74. The difference in the strength of the correlations can be partially explained by the fact that "text-based information occupies a considerable proportion of the online world". Therefore, problem-solving proficiency in technology-rich environments should be seen "in terms of proficiency in Literacy as well as in technology (OECD, 2013c). The implication of the high PS-TRE – Literacy correlation is that for those with low levels of proficiency in Literacy, it is difficult for them to score high on the PS-TRE scale, even if they do have some technology skills. As the OECD document (2013c) succinctly put it, “the digital divide may also thus reflect a Literacy divide” (p.96).

What is more, the correlation between proficiency scores in Literacy and Numeracy at the individual level for the sample is .84 -- a little bit lower than the correlation for the entire PIAAC sample (.87) (OECD, 2013c). For each country, Figures 25 and 26 display the correlation between proficiency in PS-TRE and Literacy, and PS-TRE and Numeracy, respectively. Both graphs confirm that high scores in Literacy and Numeracy “go hand in hand” with high scores in PS-TRE across the 14 selected countries. The correlation between PS-TRE and Literacy is the strongest in the three Scandinavian countries and the United Kingdom (up to .8) and the weakest in all Central European countries except the Czech Republic -- .64 in the Slovak Republic and .74 in Poland and Germany. The correlations between PS-TRE and Numeracy, although slightly smaller than that between PS-TRE and Literacy, follows a similar pattern.

Figure 25. Scatterplot of the Mean PS-TRE Scores vs. the Mean Literacy Scores across the 14 Countries

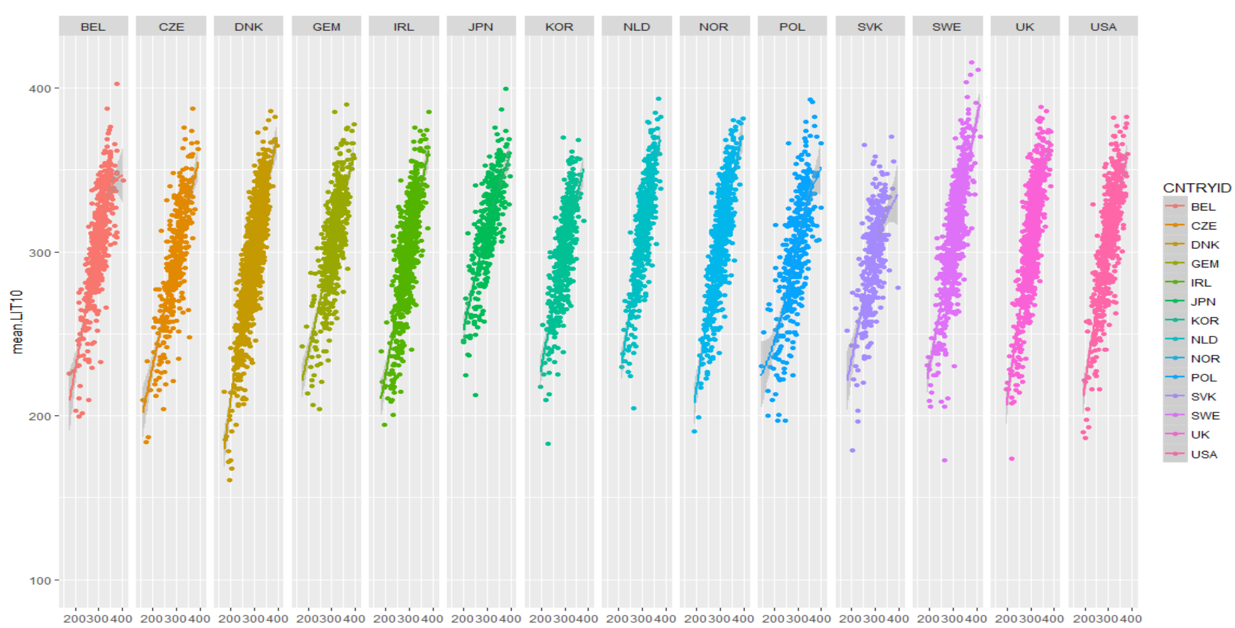
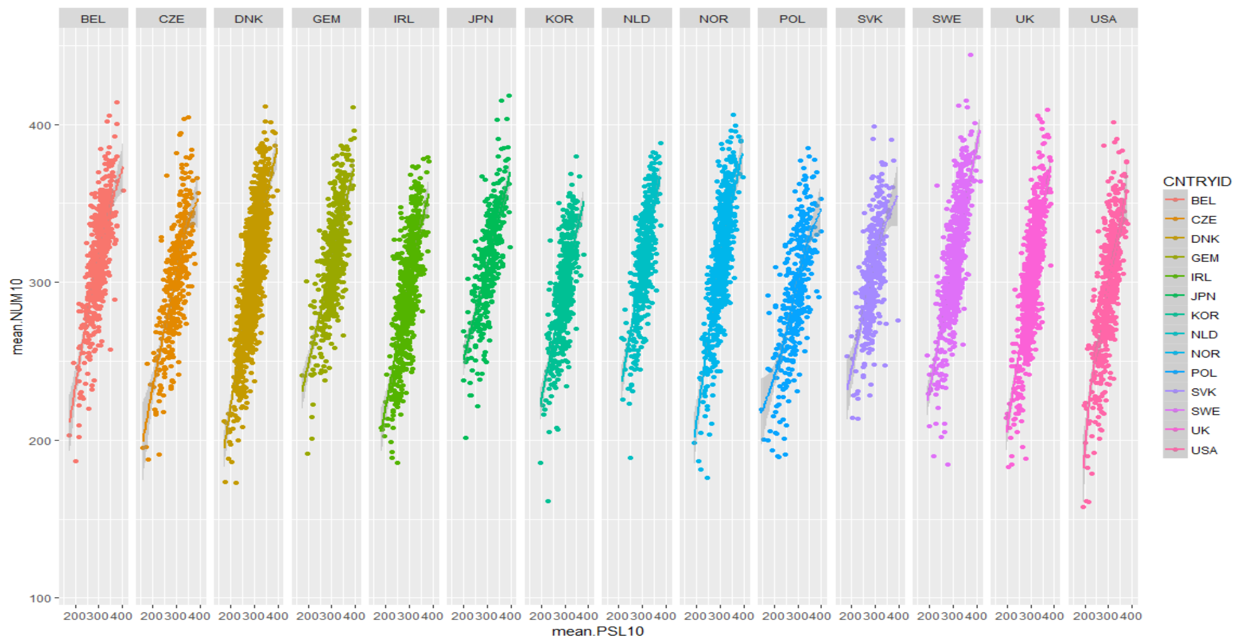


Figure 26. Scatterplot of the Mean PS-TRE Scores vs. the Mean Numeracy Scores across the 14 Countries



However, with such high correlations, the sampling distribution of Pearson's r tends to have a strong negative skew. To facilitate cross-country comparisons of the correlation between PS-TRE and the traditional cognitive skills (i.e., Literacy and Numeracy), we used Fisher's transformation to convert all the country-level correlations to approximately normally distributed z scores. As shown in Figures 27 and 28, the PS-TRE -- Literacy correlation is the highest in Denmark (1.20), the Netherlands (1.14), Norway and the United Kingdom (1.12) and lowest in the Slovak Republic (.76), Poland (.94) and Germany (.96). Similarly, the PS-TRE and Numeracy correlation is highest in Denmark (1.06), the United Kingdom (1.04) and Norway (1.02) and lowest in the Slovak Republic (.74), Poland (.88) and Germany (.89). When it comes to the relationship between the two traditional cognitive skills, the correlation is highest in Denmark (1.5) and Sweden (1.4), followed by Norway, the United Kingdom as well as the two East Asian Countries (1.3).

Figure 27. Fisher's Transformation of the Correlation between PS-TRE and Literacy

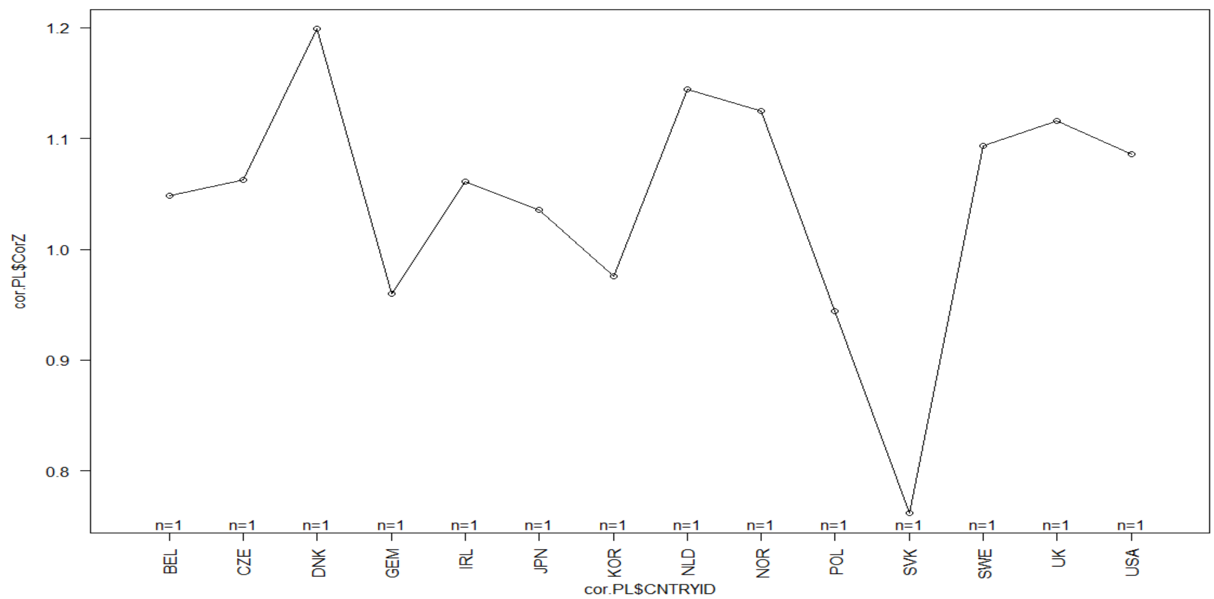
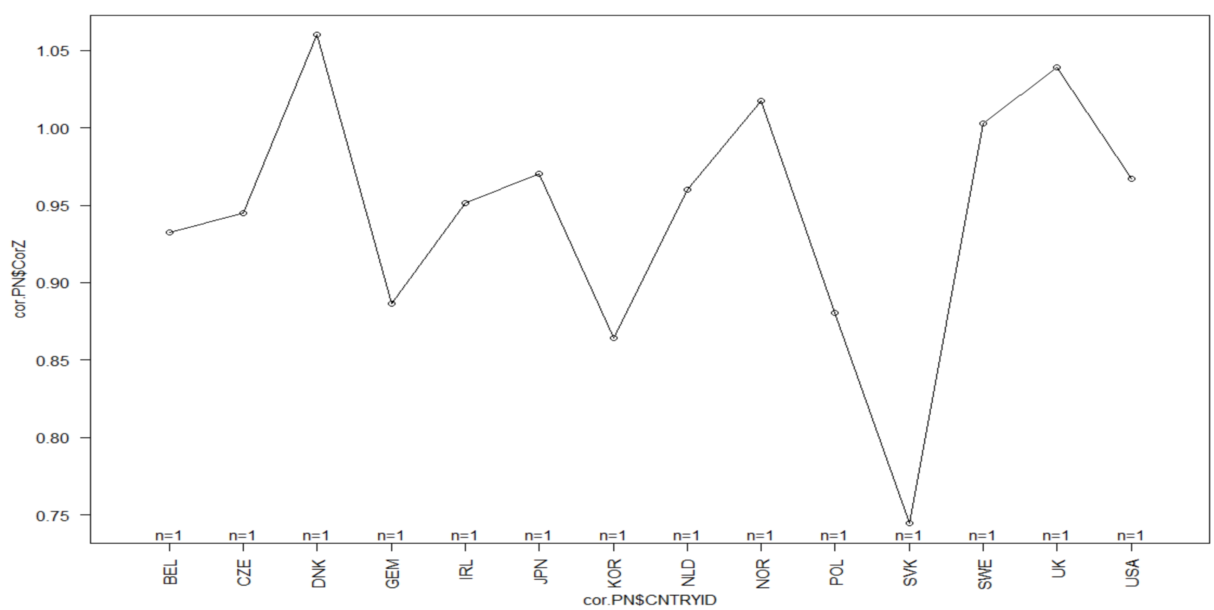


Figure 28. Fisher's Transformation of the Correlation between PS-TRE and Numeracy



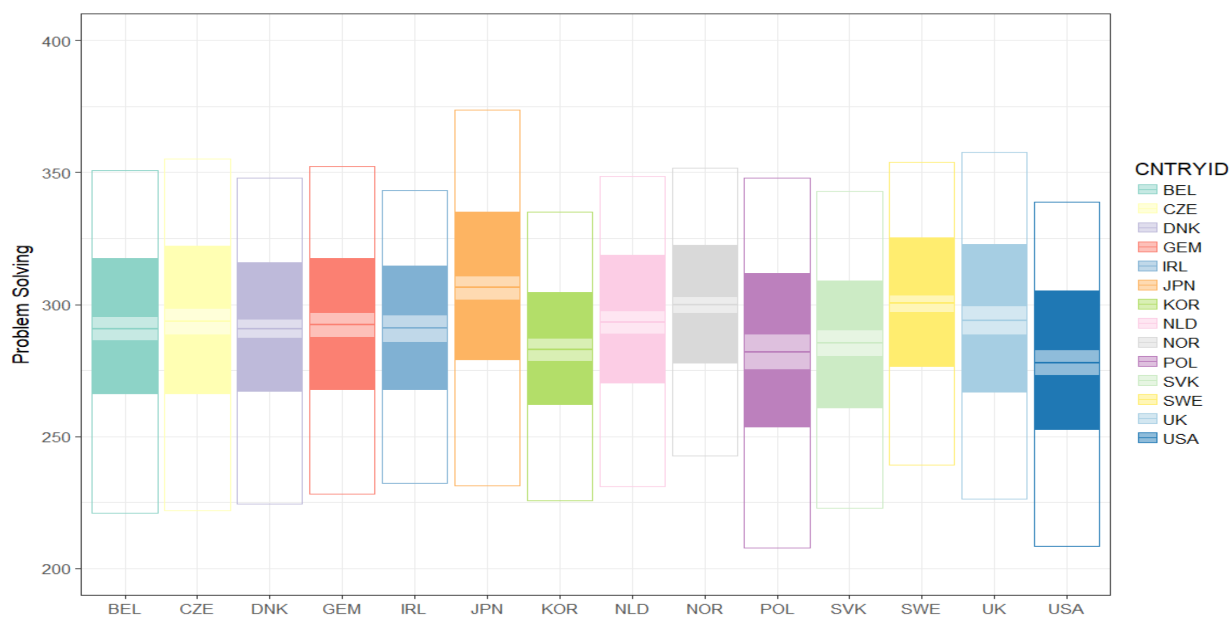
To sum up, based on the PIAAC assessment, the three key information-processing skills are most closely inter-correlated in the three Scandinavian countries and the United Kingdom, and are least strongly associated in the Central European countries. It is also interesting to note that in contrast to the relatively high Literacy-Numeracy association, the relationships between PS-TRE and the traditional cognitive skills are quite weak in South Korea and Japan (.98 and 1.04 respectively), only slightly higher than those in the aforementioned Central European countries.

Last but not least, given the differences in the correlations between PS-TRE and Literacy across countries, one-way ANOVA post-hoc tests were conducted to determine which two countries are significantly different with regard to the PS-TRE -- Literacy correlation at an alpha level of .05. Tukey's honestly significant difference (HSD) tests made pairwise comparisons between every possible pair of countries and found that there is no significant difference in the PS-TRE -- Literacy correlation between these countries, except the Slovak Republic and Denmark, and the Slovak Republic and the United Kingdom. In the same manner, it was found that there is no significant difference in the PS-TRE and Numeracy correlation between these countries, except the Slovak Republic and Denmark. Such results are not unexpected because the Slovak Republic and Denmark (and the United Kingdom) are on opposite end in terms of the strength of correlations between PS-TRE and the traditional cognitive skills. Please refer to Appendix G for the complete list of results from one-way ANOVA post-hoc tests.

4.1.4 RQ1d

RQ1d discusses how the distributions of full-time professionals' PS-TRE proficiency scores compare to those of full-time associates. As with RQ1a, Figure 29 (or Figure 21b) presents the cross-country comparisons of full-time associates' mean PS-TRE proficiency score and the 95% CI, the interquartile range (IQR) as well as the 5^{th} and 95^{th} percentiles. Table 13 (or Table 12b) ranked the 14 selected countries by the mean score calculated from the 10 plausible values -- the standard error of the mean was also considered in case of a tie. A comparison between Figure 21 and Figure 29/Figure 21b informs us that full-time professionals in all 14 countries have a higher mean PS-TRE proficiency score than full-time associates. The most dramatic difference occurs in the United States (20 points, or from 298 to 278) while the least in Ireland (1 point, or from 292 to 291). As a result, the U.S. ranking of PS-TRE proficiency dropped from the 10th in the sample of full-time professionals to the very bottom in the associated sample (see Table 13/Table 12b). On the high end, Japan has replaced Sweden as the country with the highest mean PS-TRE proficiency score, followed by two Scandinavian countries (Norway and Sweden). As a side note, all the three Scandinavian countries enjoyed the lowest standard error of the mean.

Figure 29. The Distribution of Full-time Associates PS-TRE Proficiency Scores by Country (Figure 21b, corresponding to Figure 21)



In terms of proficiency levels as defined by the OECD document (2013c), 10 of the 14 countries in the associated sample have the average proficiency at Level 2 (scores from 291 to 340) in PS-TRE, compared to all 14 countries in the sample of full-time professionals. However, average full-time associates in the Slovak Republic, South Korea, Poland and the United States performed at Level 1 where they can "complete tasks in which the goal is explicitly stated and for which the necessary operations are performed in a single and familiar environment. They can solve problems in the context of technology-rich environments whose solutions involve a relatively small number of steps, the use of a restricted range of operators, and a limited amount of monitoring across a large number of actions." (OECD, 2013c).

Across the 14 selected countries, full-time associates at the 80th percentile of the score distribution, are performing at Level 2 and Japan stands out as the only country that performed at the highest level (Level 3 with scores higher than 340) at the 80th percentile. Moving to the 85th percentile of the score distribution did not increase the number of countries performing at Level 3, compared to six countries in the sample of full-time professionals. At the 90th percentile, five countries have achieved the highest level of performance in PS-TRE, compared to 11 countries in the sample of full-time professionals.

Table 13. The Rankings of Countries by the Mean PS-TRE Proficiency Score for Full-time Associates (Table 12b, corresponding to Table 12)

CNTRYID	# of Cases	Mean Score	s.e.	80th percentile	85th percentile	90th percentile
Japan	*388*	306	2.34	**342**	**349**	**358**
Norway	*494*	300	1.58	327	333	**341**
Sweden	*476*	300	1.66	331	337	**343**
The Czech Republic	*447*	294	2.61	328	335	**343**
The United Kingdom	*484*	294	2.82	329	335	**344**
The Netherlands	*321*	293	2.25	325	332	339
Germany	*372*	292	2.43	323	329	338
Denmark	*513*	291	1.82	322	328	337
Belgium	*348*	291	2.37	323	330	339
Ireland	*254*	291	2.63	320	326	335
The Slovak Republic	*400*	285	2.55	315	322	330
South Korea	*346*	283	2.24	309	315	323
Poland	*272*	282	3.45	319	327	336
The United States	*399*	278	2.43	312	319	328

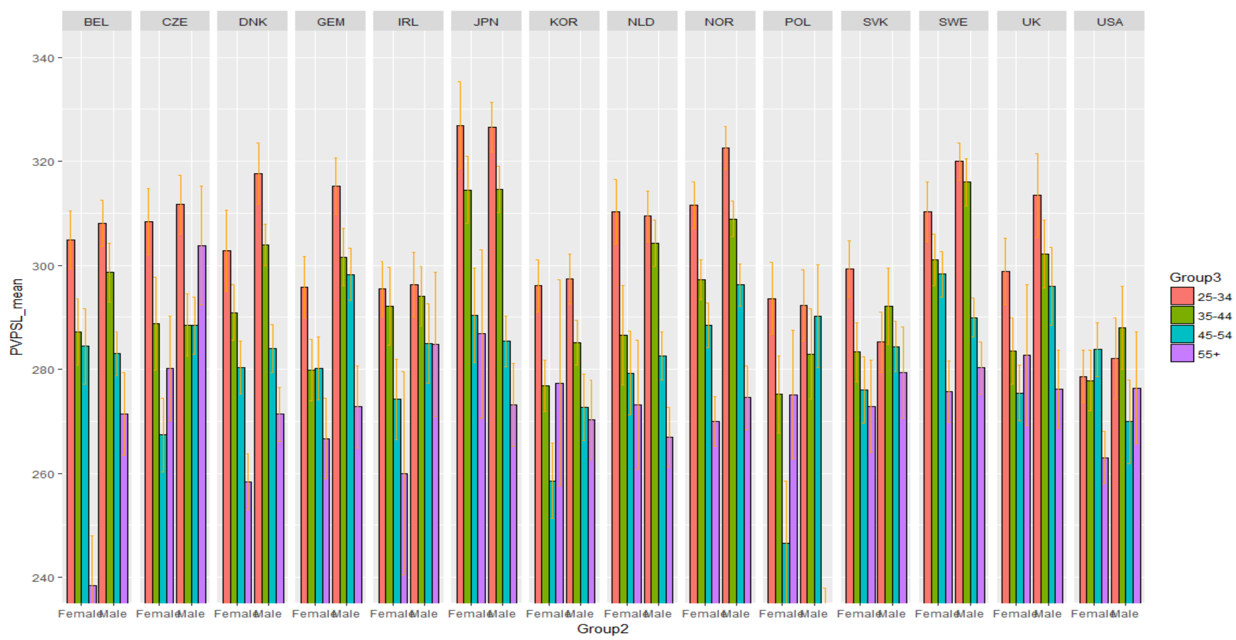
As with RQ1b, Figure 30 (or Figure 24b) illustrates the with-in country variations of mean PS-TRE proficiency scores for different combinations of Gender*Age groups. For both male and female full-time associates, Japan is the country with the highest mean PS-TRE proficiency score (305 and 311 respectively, s.e. = 2.34) while the United States became the country where the lowest mean PS-TRE proficiency score is observed (279 and 277 respectively, s.e. = 2.43). As is the case with the sample of full-time professionals, male scored higher than female full-time associates in most countries, with the exception of Japan and the Netherlands. In absolute value, the gap in favor of males is the largest in Germany (delta = 15, s.e. = 2.43) and the United Kingdom (delta = 15, s.e. = 2.82) and the smallest in the Netherlands (i.e., 1 point in favor of females, s.e. = 2.25). Comparing gender difference in PS-TRE proficiency scores between the two samples leads to two interesting findings:

(1) Japan was the country with the greatest gender gap in favor of males in the sample of full-time professionals, but reported the greatest gender gap in favor of females in the associated sample.
(2) The Czech Republic was the country with the smallest gender gap favoring males in the sample of full-time professionals; however, the gap became the second-largest gender in the associated sample.

As far as age-related differences are concerned, the youngest full-time associates (i.e., 25-34 year olds) consistently scored higher than older cohorts in all 14 countries. For the youngest cohort, the highest mean PS-TRE proficiency score was obtained by full-time associates in Japan (327) while the lowest was observed in the United States (280). Although disappointing, these findings echoed several past studies on the skills of U.S. adults in general (OECD, 2013e; Goodman et al., 2015). In other words, young people in the U.S., including those working as full-time associates, are not keeping up with their peers internationally. For the oldest cohort (i.e., 55-65 year olds), the highest mean PS-TRE proficiency score was obtained by the Czech Republic (292) and the lowest was observed in Poland (257).

As is the case with the sample of full-time professionals, the generation gap in PS-TRE proficiency is the largest in Japan (delta = 53, s.e. = 2.34) and the smallest in the United States (delta = 12, s.e. = 2.43). That is to say, young full-time associates in the United States is not doing much better than the oldest cohort of Americans. What is more worrisome, young full-time associates in all other 13 countries have made larger gains in PS-TRE performance over the oldest cohort. As scholars who wrote the OECD's PIAAC U.S. Country Note (OECD, 2013e) rightly warned: "This means that the generation of adults who will be carrying the U.S. economy for the next 30 to 40 years is entering a highly competitive global job market with almost the same skill level as those who are retiring. Without significant improvements in their skills, they will most likely be at a disadvantage to higher skilled young adults in other countries in the global competition for jobs."

Figure 30. Distribution of Full-time Associates PS-TRE Proficiency Scores by Gender, Age and Country (Figure 24b, corresponding to Figure 24)



When splitting the associated sample by both gender and age groups, the age-related declines persisted in most countries, except in the three Visegrád Group countries and the United States. In the Czech Republic, the oldest male full-time associates outperformed their younger peers aged between 35-54 by 16 points (s.e. = 2.61); while in Poland, the oldest female full-time associates outperformed their 45-54-year-old peers by 28 points (s.e. = 3.45). In the Slovak Republic, male full-time associates aged 35-44 outperformed the youngest male group by 7 points (s.e. = 2.55). Again, it is worrisome to find out that the United States is the only country where the youngest cohort did not have the highest score with either gender; instead, 35-44-year-old male and 45-54-year-old female full-time associates have the highest mean PS-TRE proficiency score respectively.

As with RQ1c, we examined the correlations between full-time associates' PS-TRE proficiency scores with their proficiency scores in Literacy and Numeracy. The results are very close to that of full-time professionals. On average across the 14 selected countries, the correlation between proficiency scores in PS-TRE and Literacy is .80 (compared with .79 for full-time professionals), while the correlation between PS-TRE and Numeracy is .75 (compared with .74 for full-time professionals). And the correlation between Literacy and Numeracy proficiencies is .80, slightly lower than .84 in the sample of full-time professionals.

To facilitate cross-country comparisons, Figures 31 and 32 show the standardized correlation (i.e, Fisher's Z) between PS-TRE and the two traditional cognitive skills respectively. Again, the results are highly similar to that of full-time professionals, with the exception of Sweden. Specifically, Sweden is the only Scandinavian country with both inter-correlations located to the lower end of the range.

Figure 31. Fisher's Transformation of the Correlation between PS-TRE and Literacy (Figure 27b, corresponding to Figure 27)

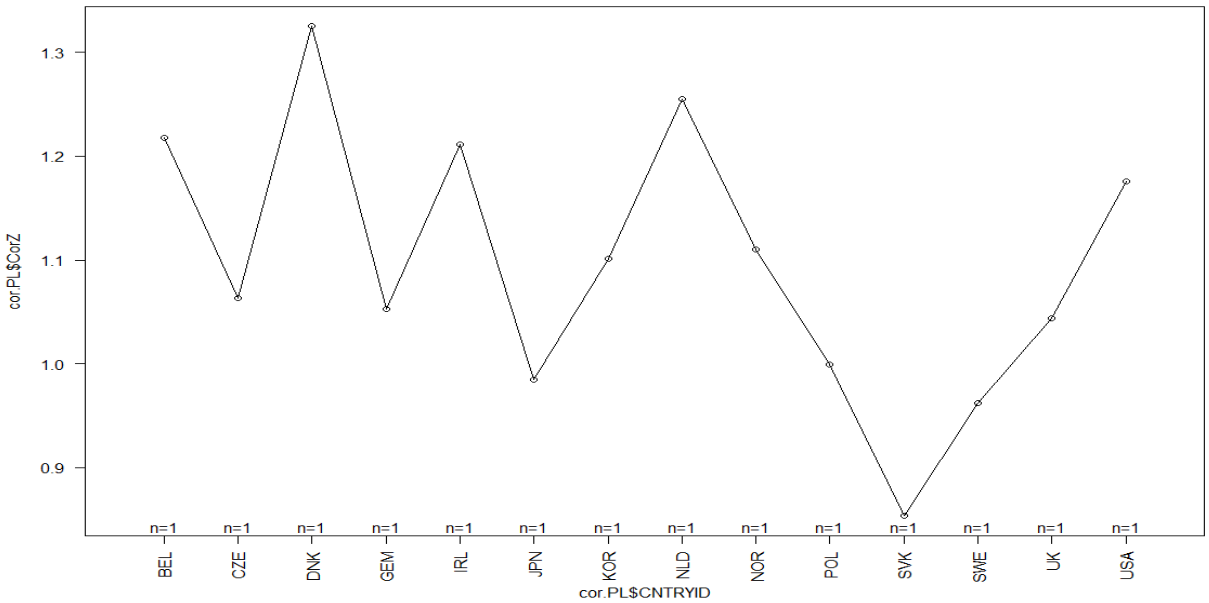
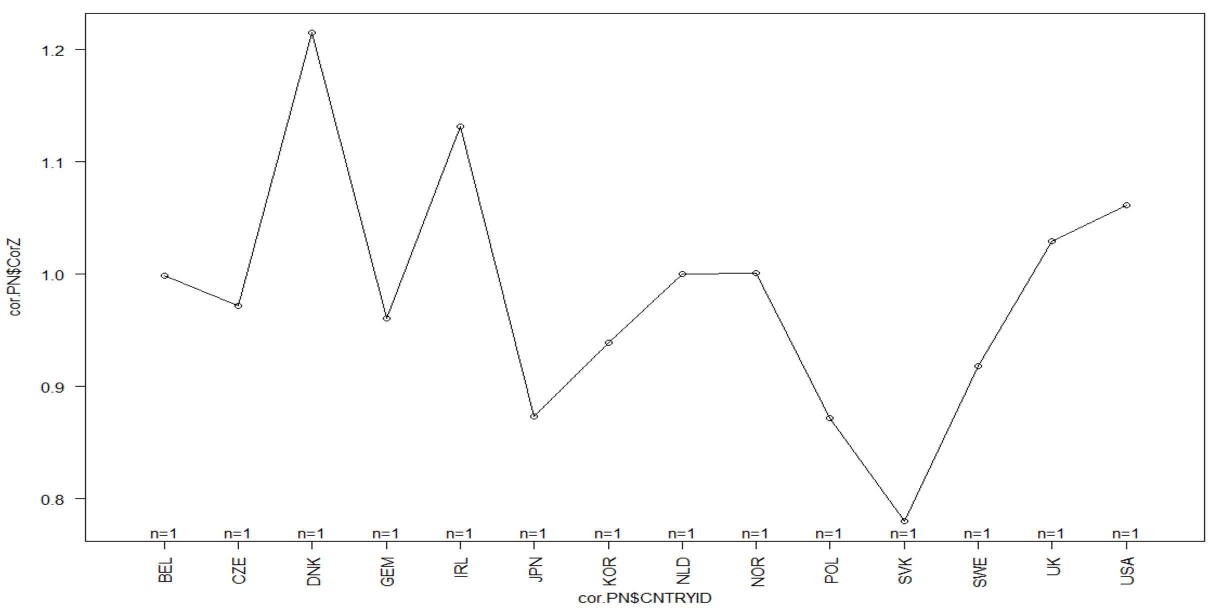


Figure 32. Fisher's Transformation of the Correlation between PS-TRE and Numeracy (Figure 28b, corresponding to Figure 28)

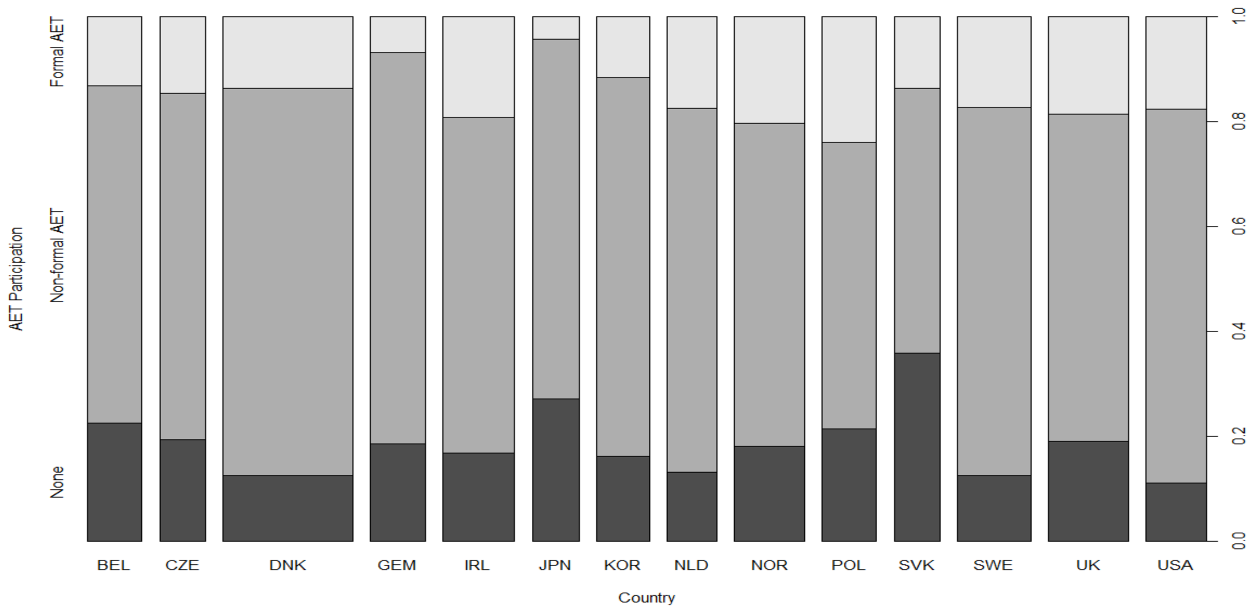


Different from the sample of full-time professionals where there were significant between-country differences in correlations between PS-TRE and Literacy and Numeracy respectively, one-way ANOVA post-hoc tests found no significant difference in correlations between PS-TRE and Literacy or Numeracy across the 14 selected countries (p-value equals .24 and .09 respectively).

4.2 Research Question Two

Research question two focuses on predicting full-time professionals' probabilities of participating in the different formats of AET programs, based on four clusters of predictors: their socio-demographic background factors, occupational categories, and the types and intensities of skills use at work (and at home). As introduced in Section 3.4.3, there are three formats of AET participation -- Formal, Non-formal and None. Figure 33 provides an overview of full-time professionals' participation in the three formats of AET programs by country. It is evident that the 14 selected countries differ substantially not only in the distribution of participation rates across the three formats of AET programs, but also in the sizes of the sampled full-time professionals (N = 8,535). Therefore, it is meaningful to build predictive models for each country.

Figure 33. Overview of Full-time Professionals' Participation in the Different Formats of AET Programs by Country



As introduced in Section 3.3, Multinomial logistic regression is actually a multi-equation model. For the nominal dependent variable with three categories, the multinomial logistic regression model estimates 3-1 = 2 logit equations. With None AET participation as the reference category, the two logit equations estimate the log odds of Non-formal AET participation vs. None and Formal AET Participation vs. None respectively. For each logit equation, the above-named four clusters of predictor variables are entered stepwise to specify the model. Section 2.1 explains the rationale behind predictor selection and Table 6 in Section 3.2 provides an overview of the 17 selected predictors by cluster. Because the model specification of multinomial logistic regression strongly resembles that of standard multiple regression, we will not go into detail about the model-building process, but rather focus on understanding the results of multinomial logistic regression models.

4.2.1 Non-Formal AET Participation vs. None

Given the large amount of information generated as outputs for the multinomial logistic regression procedure, it would be helpful to present results from the aforementioned two logit equations separately. Specifically, Sections 4.2.1 and 4.2.2 report parameter estimations regarding the log odds of Non-formal AET participation vs. None and Formal AET participation vs. None respectively. The focus will be on classifying the 14 selected countries based on the clustering of significant predictors at the .01 level.

Group 1. Countries where occupational categories significantly predict Non-formal AET participation – Belgium, the Czech Republic, Ireland and the Slovak Republic

Tables 14 -27 below shows significant coefficients of the multinomial logistic regression model for Non-formal AET Participation vs. None for each of the 14 selected countries. Recall that a categorical predictor with k levels is represented by k-1 binary indicators in the model (i.e., binary coding). And each binary indicator has a coefficient indicating the comparison between this specific level and the reference level which, by default of SPSS Statistics, is the last numeric level of the given predictor.

Table 14. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Belgium

	B	Std. Error	Sig.	Exp(B)
Sci&Eng	1.81	.59	.002	6.12
Health	1.49	.47	.001	4.43

* *Reference Category: Legal, Social and Cultural Professionals*

In Belgium, the only significant predictors for full-time professionals' Non-formal AET Participation (vs. None) are two occupational categories – Science & Engineering and Health. That is to say, after partialling out the variability due to the other variables in the model, the odds of Non-formal AET participation for Belgian professionals in these two categories are higher than the odds for those in Legal, Social and Cultural (i.e. the reference level).

Table 15. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – the Slovak Republic

	B	Std. Error	Sig.	Exp(B)
Sci&Eng	1.21	.40	.003	3.35
Health	2.13	.57	.000	8.37
EDU less than B.A.	-.81	.27	.002	.45

* *Reference Category for Occupational Categories: Legal, Social and Cultural Professionals*
* *Reference Category for Educational Attainment: A Master's degree or above*

As can be seen from Table 15, the Slovak Republic is highly similar to Belgium in terms of the occupational categories that significantly predict full-time professionals' Non-formal AET participation (vs. None). Moreover, for two individuals with the same predictor profile except for Education Attainment, the odds of Non-formal AET participation for those with less than a Bachelor's degree are lower than those with a Master's degree or above.

Table 16. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – the Czech Republic

	B	Std. Error	Sig.	Exp(B)
Biz&Admin	1.65	.63	.009	5.19

* *Reference Category: Legal, Social and Cultural Professionals*

Table 17. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Ireland

	B	Std. Error	Sig.	Exp(B)
NUMHOME	.38	.15	.011	1.46
Teaching	1.20	.48	.013	3.31

* *Reference Category: Legal, Social and Cultural Professionals*

In the Czech Republic, Business & Administration is the only significant predictors for full-time professionals' Non-formal AET Participation (vs. None). In Ireland, it is Teaching that significantly predicts full-time professionals' Non-formal AET participation (vs. None).

Group 2. Countries where educational attainment significantly predicts Non-formal AET participation – Japan, South Korea

Educational attainment plays a significant role in determining whether full-time professionals will participate in non-formal AET programs (vs. None) in the two East Asian countries. As shown in Tables 18 and 19 below, Japan and South Korea are similar in terms of the magnitudes of the coefficients corresponding to Education Attainment. That is tosay, for two Japanese full-time professionals with the same predictor profile except for Education Attainment, a higher degree level is significantly associated with an increase in the odds of participating in Non-formal AET participation. The relationship between full-time professionals' Education Attainment and Non-formal AET participation is even stronger among full-time professionals in South Korea.

Table 18. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Japan

	B	Std. Error	Sig.	Exp(B)
LEARN@WORK	.67	.20	.001	6.12
EDU less than B.A.	-1.77	.57	.002	4.43
EDU B.A. only	-1.35	.53	.010	.26

** Reference Category for Educational Attainment: A Master's degree or above*

Table 19. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – South Korea

	B	Std. Error	Sig.	Exp(B)
EDU less than B.A.	-2.43	.66	.000	.09
EDU B.A. only	-1.58	.65	.015	.21

** Reference Category for Educational Attainment: A Master's degree or above*

In addition to Education Attainment, full-time professionals' use of skills at work can also be related to Non-formal AET participation. In Japan, after partialling out the variability due to the other variables in the model, a higher intensity of using Learning at Work skills is significantly associated with an increase in the odds of participating in Non-formal AET programs.

Group 3. Countries where the use of skills significantly predicts Non-formal AET participation – Denmark, Germany, the Netherlands, and the United Kingdom

A careful examination of parameter estimates across the 14 selected countries suggests that after partialling out the variability due to the other variables in the model, the self-reported intensity of skills use serves as a significant predictor in as many as six countries. In Denmark, Germany and the Netherlands, full-time professionals' Non-formal AET participation (vs. None) is only significantly predicted by a single indicator of skill use at work.

Table 20. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Denmark

	B	Std. Error	Sig.	Exp(B)
READWORK	.65	.19	.001	1.90

Table 21. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – Germany

	B	Std. Error	Sig.	Exp(B)
READWORK	.56	.22	.009	1.75

Table 22. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – The Netherlands

	B	Std. Error	Sig.	Exp(B)
NUMWORK	-.50	.18	.005	.61

Table 23. Significant Predictors for Full-time Professionals' Non-formal AET Participation vs. None – The United Kingdom

	B	Std. Error	Sig.	Exp(B)
WRITEWORK	.58	.17	.001	1.78
INFLUENCE	.41	.16	.011	1.50
EDU less than B.A.	-.67	.27	.012	.51

* *Reference Category for Educational Attainment: A Master's degree or above*

Table 24 below summarizes the distribution of significant indicators of skills use across these countries at the .01 level. After partialling out the variability due to the other variables in the model, the use of Reading skills at work is significantly associated with an increase in the odds of Non-formal AET participation in Denmark and Germany while the use of Writing skills at work is significantly associated with an increase in the odds of Non-formal AET participation in the United Kingdom. What is more, for two British full-time professionals with the same predictor profile except for the use of Influence skills at work, a one-unit increase in the predictor Influence is significantly associated with 50% increase in the odds of Non-formal AET participation.

It is worth noting that with after partialling out the variability due to the other variables in the model, a one-unit increase in the use of Numeracy skills at work is significantly associated with 40% decrease in the odds of Non-formal AET participation in the Netherlands while a one-unit increase in the use of Numeracy skills at home is significantly associated with 50% increase in the odds of Non-formal AET participation in Ireland. Somewhat surprisingly, the use of ICT skills (either at work or at home) appears to be unrelated to the odds of Non-formal AET participation in any of the 14 selected countries.

Table 24. Distribution of Significant Indicators of Skills Use across Countries – Non-formal vs. None

	Indicator of Skills Use	**Country (with Odds Ratio)**
Use of Key Information-Processing Skills at Work	READWORK	Denmark (1.9), Germany (1.7)
	WRITEWORK	The United Kingdom (1.8)
	NUMWORK	*The Netherlands (0.6)*
Use of Key Information-Processing Skills at Home	NUMHOME	Ireland (1.5)
Use of Generic Workplace Skills/Job-related Activities	INFLUENCE	The United Kingdom (1.5)
	LEARATWORK	Japan (1.9)

Group 4. Countries with no significant predictor for Non-formal AET participation – Norway, Poland, Sweden and the United States

Lastly, there are up to four countries where no significant predictor of full-time professionals' Non-formal AET participation was found. The implication is that the specified multinomial logistic model did fit the sample from these countries. For the next section, we are curious to find out if such classification of countries applies to the prediction of full-time professionals' Formal AET participation (vs. None).

4.2.2 Formal AET Participation vs. None

Group 1. Countries where age cohorts significantly predict Formal AET participation – Belgium, the Czech Republic, Denmark, Sweden and the United States

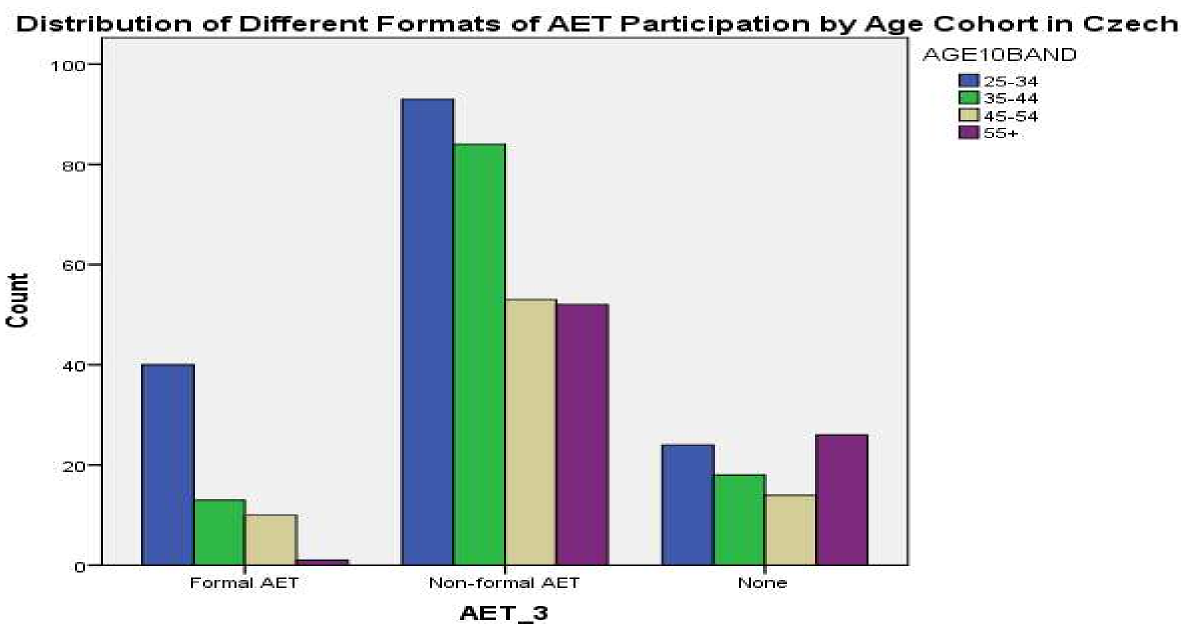
In contrast to the previous section where age cohorts appear to be unrelated to the odds of Non-formal AET participation in any of the 14 selected countries, age cohorts turn out to be the most important predictors of full-time professionals' Formal AET participation in five countries. In the Czech Republic, all age cohorts younger than 55 year olds are significantly associated with an increase in the odds of Formal AET participation, after partialling out the variability due to the other variables in the model. However, an examination of the distribution of different formats of AET participation by age cohort reveals that there are only one 55-65 year olds who participated in Formal AET programs in the Czech Republic, therefore, the Czech results will not be considered for later discussion.

Table 25. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Czech Republic

	B	Std. Error	Sig.	Exp(B)
LEARN@WORK	.95	.29	.001	2.57
AGE 25-34	3.91	1.13	.001	49.71
AGE 35-44	3.42	1.16	.003	30.58
AGE 45-54	3.02	1.18	.010	20.52

** Reference Category for Age in 10-year bands: 55-65 year olds*

Figure 34. Distribution of Different Formats of AET Participation by Age Cohort in Czech



The strong association between age cohorts and full-time professionals' Formal AET participation is also observed in Denmark where all age cohorts younger than 45 year olds are significantly associated with an increase in the odds of Formal AET participation, after partialling out the variability due to the other variables in the model. What is more, after partialling out the variability due to the other variables in the model, the odds of participating in Formal AET programs for those between 25 and 34 years old are significantly higher than those between 55 and 65 years old in Belgium, Sweden and the United States.

Table 26. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Denmark

	B	Std. Error	Sig.	Exp(B)
READWORK	.81	.24	.001	2.25
AGE 25-34	2.58	.43	.000	13.25
AGE 35-44	1.13	.37	.002	3.10

** Reference Category for Age in 10-year bands: 55-65 year olds*

Table 27. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Belgium

	B	Std. Error	Sig.	Exp(B)
Health	1.96	.70	.005	7.10
AGE 25-34	2.62	.89	.003	13.67

** Reference Category for Occupational Categories: Legal, Social and Cultural Professionals*
** Reference Category for Age in 10-year bands: 55-65 year olds*

Table 28. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Sweden

	B	Std. Error	Sig.	Exp(B)
READWORK	.83	.34	.014	2.30
AGE 25-34	1.93	.53	.000	6.91

** Reference Category for Age in 10-year bands: 55-65 year olds*

Table 29. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the United States

	B	Std. Error	Sig.	Exp(B)
AGE 25-34	1.84	.59	.002	6.29

** Reference Category for Age in 10-year bands: 55-65 year olds*

Group 2. Countries where educational attainment significantly predicts Non-formal AET participation – Japan and South Korea

Similar to Group 2 in the previous section for Non-formal AET participation, Education Attainment is key in predicting full-time professionals' Formal AET participation in the two East Asian Countries. As shown in Tables 30 and 31, for two Japanese full-time professionals with the same predictor profile except for Education Attainment, a higher degree level is significantly associated with an increase in the odds of participating in Formal AET programs, and such relationship is even stronger among full-time professionals in South Korea.

Table 30. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Japan

	B	Std. Error	Sig.	Exp(B)
EDU less than B.A.	-2.76	1.04	.008	.06
EDU B.A. only	-2.47	.85	.004	.09

** Reference Category for Educational Attainment: A Master's degree or above*

Table 31. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – South Korea

	B	Std. Error	Sig.	Exp(B)
EDU less than B.A.	-3.42	.78	.000	.033
EDU B.A. only	-2.62	.74	.000	.073

** Reference Category for Educational Attainment: A Master's degree or above*

Group 3. Countries where occupational categories significantly predict Formal AET participation – Belgium and the Slovak Republic

Occupational categories also play a role in predicting full-time professionals' Formal AET participation (vs. None) in three countries. Recall that in previous section, Science &

Engineering and Health are the only two significant predictors for professionals' Non-formal AET participation in Belgium. When it comes to Formal AET participation (vs. None), the odds for Health professionals are still significantly higher than the odds for Legal, Social and Cultural professionals. Please refer to Table 27 above for more details.

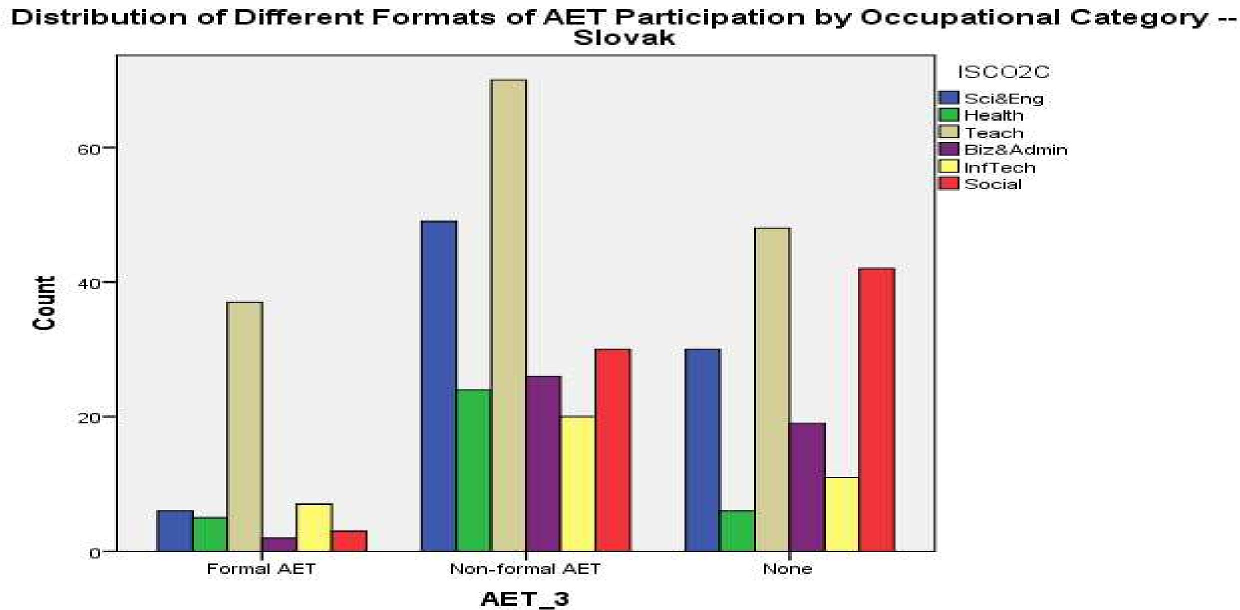
Again, the Slovak Republic is similar to Belgium in terms of the occupational categories that significantly predict full-time professionals' Formal AET participation (vs. None). However, due to the unbalanced distribution of occupational categories among Formal AET participants (see Figure 35 below), the Slovak results will be excluded considered from later discussion.

Table 32. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Slovak Republic

	B	Std. Error	Sig.	Exp(B)
Health	2.96	.93	.001	19.27
Teaching	3.22	.76	.000	25.03
InfoTech	2.35	.86	.007	10.45

** Reference Category for Occupational Categories: Legal, Social and Cultural Professionals*

Figure 35. Distribution of Different Formats of AET Participation by Occupational Category in the Slovak Republic



Group 4. Countries where the use of skills significantly predicts Formal AET participation – the Czech Republic, the Netherlands, Norway and the United Kingdom

The self-reported intensity of skills use significantly predicts full-time professionals' Formal AET participation in as many as six countries. Among them, two belong to Group 1 where age cohorts have a stronger relationship with Formal AET participation – Denmark and Sweden (see Tables 26 and 28). In the rest four countries -- Czech Republic, the Netherlands, Norway and the United Kingdom -- indicators of skill use are the only significant predictors. Please see Tables 33-36 below for more details.

Table 33. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Czech Republic

	B	Std. Error	Sig.	Exp(B)
LEARN@WORK	.95	.29	.001	2.58

Table 34. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the Netherlands

	B	Std. Error	Sig.	Exp(B)
WRITEWORK	.83	.32	.010	2.30

Table 35. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – Norway

	B	Std. Error	Sig.	Exp(B)
NUMHOME	.69	.24	.003	2.00

Table 36. Significant Predictors for Full-time Professionals' Formal AET Participation vs. None – the United Kingdom

	B	Std. Error	Sig.	Exp(B)
INFLUENCE	.46	.19	.014	1.58
TASKDISC	-.45	.16	.006	.64

Table 37 below provides a quick overview of significant indicators of skills use across these countries at the .01 level. For full-time professionals with the same predictor profile except for the use of Reading skills at work, a one-unit increase in the predictor Reading at Work is associated with 130% increase in the odds of participating in AET programs in Denmark and

Sweden. A comparison between Table 24 and Table 37 shows that in Denmark, a higher intensity of using Reading at Work skills is significantly associated with an increase in the odds of participating in either format of AET programs, but more so for Formal AET programs – odds ratio equals 1.9 for Non-formal vs. 2.3 for Formal AET participation. A similar pattern is observed in the United Kingdom where a one-unit increase in the use of Influence skills at work is associated with an increase in the odds of participating in both formats of AET programs – 50% increase for Non-formal vs. 60% increase for Formal AET participation. On the other hand, a one-unit increase in the use of Task Discretion skills at Work is associated with 36% decrease in the odds of Formal AET participation in the United Kingdom, after partialling out the variability due to the other variables in the model.

As to the use of key information-processing skills at home, Numeracy appears to be the only significant predictor for full-time professionals' AET participation (either Formal or Non-formal). A one-unit increase in the use of Numeracy skills at home is associated with 46% increase in the odds of Non-formal AET participation in Ireland and 100% increase in the odds of Formal AET participation in Norway. Please refer to Tables 24 and 37 for more details.

Table 37. Distribution of Significant Indicators of Skills Use across Countries – Formal vs. None AET Participation

	Indicator of Skills Use	Country (Odds Ratio)
Use of Key Information-Processing Skills at Work	READWORK	Denmark (2.3) Sweden (2.3)
	WRITEWORK	The Netherlands (2.3)
Use of Key Information-Processing Skills at Home	NUMHOME	Norway (2.0)
Use of Generic Workplace Skills/Job-related Activities	INFLUENCE	The United Kingdom (1.6)
	TASKDISC	*The United Kingdom (.64)*
	LEARATWORK	The Czech Republic (2.6)

Group 5. Countries with no significant predictor for Formal AET participation – Germany, Ireland and Poland

Recall that for full-time professionals' Non-formal AET participation, there are four countries where no significant predictor was found – Norway, Poland, Sweden and the United States. When it comes to Formal AET participation, Poland remains the country with no significant predictor, along with Germany and Ireland. Again, this lack of fit suggests potential misspecifications of the multinomial logistic model in these countries. Research question three will first evaluate the performance of the specified model (namely the propensity scores model) and then discuss methods appropriate for addressing the question at hand.

4.3 Research Question Three

4.3.1 Estimate Generalized Propensity Scores with Multinomial Logistic Regression

4.3.1.1 Predicted Probabilities for the Three Formats of AET Participation

In addition to parameter estimation, multinomial logistic regression can be used to predict probabilities for the three formats of AET participation, i.e., the generalized propensity scores (GPS). As introduced by formula 11 in Chapter 2, a baseline-category logit model is used to estimate the probability of full-time professionals' participation in one format of AET programs instead of None (i.e., the reference condition), for Non-formal and Formal AET participation respectively. Alternatively, this model can be expressed in terms of the probabilities of participating in each format of AET programs (including None AET participation). Please refer to formula 12 for the probability of participating in Non-formal/Formal AET programs and formula 13 for the probability of None AET participation. Note that the final full sample weight from PIAAC (i.e., SPFWT0) is applied to the estimation of the GPS to account for unequal probabilities of selection.

For each of the 14 selected countries, the fitted values consist of an N x 3 matrix where the probability of participation for N respondents is provided for each of the three formats of AET programs – Non-formal, Formal and None. Table 38 provides an array of such matrices that summarizes the estimated probabilities by full-time professionals' actual AET participation across countries. Each row represents the actual participants of one format of AET programs in a country whereas each column represents the estimated probability for one format of AET participation. For each row, the three propensity scores will sum to one. That is to say, for participants of the same format of AET programs in a given country, the estimated GPS for the three formats of AET participation will sum to one.

It is evident from Table 38, the probabilities of Non-formal AET participation are the highest across all three types of AET participants and the 14 selected countries. This is largely due to the fact that up to two thirds of full-time professionals in the sample reported having participated in Non-formal AET programs in the 12 months prior to the survey. Unsurprisingly, the highest probabilities of Non-formal AET participation are observed among those who actually participated in Non-formal AET programs across countries. At the country level, the highest propensity scores for Non-formal AET participation among actual Non-formal AET participants are observed in Germany (.79), the United States (.76) and Denmark (.75), whereas the lowest is observed in the Slovak Republic (.58). Also note that among those who actually participated in None AET programs, the distribution of propensity scores for Non-formal AET participation followed a similar pattern – the highest in the United States (.72), Germany (.71) and Denmark (.71), whilst the lowest in the Slovak Republic (.44). Among those who actually participated in Formal AET programs, however, the highest GPS for Non-formal AET participation are observed in Japan (.72) while the lowest is still observed in the Slovak Republic. That is to say, regardless of their actual AET participation, full-time professionals in the Slovak Republic are estimated to have the lowest GPS for Non-formal AET participation in the sample.

For those who actually participated in None AET program, the probabilities of None AET participation are the second-highest in all 14 countries – lower than that of Non-formal AET participation but higher than that of Formal AET participation. The highest propensity scores for None AET participation among actual None AET participants are observed in Japan (.38) and the three Visegrád Group countries (.46 in the Slovak Republic, .34 in the Czech Republic and .32 in Poland), whereas the lowest is observed in the United States (.17) and the three

Scandinavian countries (.18 in Denmark, .19 in Sweden and .22 in Norway). Similarly, for those who actually participated in Formal AET programs, the probabilities of Formal AET participation are the second-highest in most countries, with the exception of Japan and the Slovak Republic where Formal AET participants are estimated to have higher propensity scores for None AET participation than for Formal AET participation. This signals lack of fit of this specified model for the prediction of Formal AET participation among Japanese and Slovak full-time professionals. As a next step, we will obtain measures of goodness of fit and predictive power for multinomial logistic regression models and compare results across countries.

Table 38. Summary of the Estimated GPS Matrix for Full-time Professionals by Their Actual AET Participation by Country

BEL	**None**	**Non-formal**	**Formal**
None Participants	**.28**	.63	.09
Non-formal Participants	.21	**.67**	.12
Formal Participants	.17	.60	**.23**

CZE	**None**	**Non-formal**	**Formal**
None Participants	**.34**	.59	.07
Non-formal Participants	.20	**.69**	.11
Formal Participants	.15	.51	**.34**

DNK	**None**	**Non-formal**	**Formal**
None Participants	**.18**	.71	.11
Non-formal Participants	.12	**.75**	.13
Formal Participants	.09	.68	**.23**

GEM	**None**	**Non-formal**	**Formal**
None Participants	**.23**	.71	.06
Non-formal Participants	.16	**.79**	.05
Formal Participants	.15	.61	**.24**

IRL	**None**	**Non-formal**	**Formal**
None Participants	**.24**	.57	.19
Non-formal Participants	.15	**.67**	.18
Formal Participants	.16	.57	**.27**

JPN	**None**	**Non-formal**	**Formal**
None Participants	**.38**	.60	.02
Non-formal Participants	.22	**.74**	.03
Formal Participants	.17	.72	**.11**

KOR	**None**	**Non-formal**	**Formal**
None Participants	**.31**	.63	.06
Non-formal Participants	.15	**.75**	.10
Formal Participants	.10	.65	**.25**

NLD	**None**	**Non-formal**	**Formal**
None Participants	**.26**	.64	.10
Non-formal Participants	.12	**.73**	.15
Formal Participants	.08	.64	**.28**

NOR	**None**	**Non-formal**	**Formal**
None Participants	**.22**	.62	.17
Non-formal Participants	.17	**.65**	.18
Formal Participants	.15	.55	**.30**

POL	**None**	**Non-formal**	**Formal**
None Participants	**.32**	.51	.17
Non-formal Participants	.19	**.64**	.16
Formal Participants	.18	.53	**.29**

SVK	**None**	**Non-formal**	**Formal**
None Participants	**.46**	.44	.10
Non-formal Participants	.30	**.58**	.12
Formal Participants	.27	.48	**.26**

SWE	**None**	**Non-formal**	**Formal**
None Participants	**.19**	.68	.14
Non-formal Participants	.12	**.73**	.14
Formal Participants	.11	.56	**.34**

UK	**None**	**Non-formal**	**Formal**
None Participants	**.30**	.57	.13
Non-formal Participants	.15	**.68**	.17
Formal Participants	.14	.62	**.24**

USA	**None**	**Non-formal**	**Formal**
None Participants	**.17**	.72	.11
Non-formal Participants	.11	**.76**	.13
Formal Participants	.08	.62	**.30**

4.3.1.2 Measures of Goodness of Fit and Predictive Power

After fitting multinomial logistic regression models, it is important to assess how well the specified model fits the data at the country level. Approaches to answering this question generally fall into two categories – goodness of fit tests (e.g., Pearson chi-square) and measures of predictive power (e.g., Pseudo R-square) (Allison, 2014). In this section, we will look at both measures of model fit and make comparisons across the 14 selected countries.

Measures of Goodness of Fit

To test the null hypothesis that the specified model fits the data well, two goodness-of-fit statistics are computed. Table 39 presents the goodness-of-fit tests based on the deviance and the Pearson chi-square. The deviance, or -2 log-likelihood (-2LL) statistic, measures the deviance of the fitted model with respect to a saturated model. More precisely, the deviance is defined as the difference of the log-likelihoods between the fitted model and the saturated model. Since the log-likelihood of the fitted model is always smaller than or equal to the log-likelihood of the saturated model, the deviance is always greater than or equal to zero (being zero only if the fit of the model is perfect). If the model fits the data well, the deviance will have approximately a chi-square distribution on N – p degrees of freedom, where N is the sample size and p is the number of parameters. However, it is important to note that the goodness-of-fit test based on deviance has little intuitive meaning because it depends on the sample size and the number of parameters in the model as well as on the goodness of fit. Therefore, the extreme p-values in Table 39 should not be over-interpreted.

Alternatively, the Pearson chi-square statistic compares observed values to those predicted by the specified model and approximates a chi-square distribution (df = N – p) if the model fits the data well. You can see from Table 39 below that the Pearson chi-squares are low relative to the degrees of freedom and the p-values are greater than .01 in all but three countries -- the Czech Republic, Sweden and the United Kingdom. That is to say, based on the Pearson chi-square statistic, there is no evidence to reject the null hypothesis that the specified model fit the data well in all but three countries. Again, it is important to note that results from the Pearson chi-square test should not be over-interpreted because the approximation of chi-square distribution keeps improving as the sample gets larger (Allison, 2014).

Table 39. Goodness of Fit of Multinomial Logistic Regression Models by Country

BEL	Chi-Square	df	Sig.
Pearson	1079.83	976	.011
Deviance	820.34	976	1.000

CZE	Chi-Square	df	Sig.
Pearson	1235.85	804	.000
Deviance	615.67	804	1.000

DNK	Chi-Square	df	Sig.
Pearson	2466.51	2424	.269
Deviance	1692.44	2424	1.000

GEM	Chi-Square	df	Sig.
Pearson	899.56	974	.957
Deviance	621.48	974	1.000

IRL	Chi-Square	df	Sig.
Pearson	1400.42	1300	.027
Deviance	1110.72	1300	1.000

JPN	Chi-Square	df	Sig.
Pearson	727.49	808	.980
Deviance	543.83	808	1.000

KOR	Chi-Square	df	Sig.
Pearson	954.39	964	.581
Deviance	651.88	964	1.000

NLD	Chi-Square	df	Sig.
Pearson	942.25	890	.109
Deviance	665.31	890	1.000

NOR	Chi-Square	df	Sig.
Pearson	1309.01	1284	.307
Deviance	1135.27	1284	.999

POL	Chi-Square	df	Sig.
Pearson	1002.17	974	.259
Deviance	847.18	974	1.000

SVK	Chi-Square	df	Sig.
Pearson	811.23	818	.560
Deviance	723.40	818	.992

SWE	Chi-Square	df	Sig.
Pearson	1480.96	1326	.002
Deviance	965.95	1326	1.000

UK	Chi-Square	df	Sig.
Pearson	1479.22	1352	.009
Deviance	1153.29	1352	1.000

USA	Chi-Square	df	Sig.
Pearson	1124.08	1094	.257
Deviance	771.28	1094	1.000

Another option to get an overall measure of goodness of fit is to compare the specified model to the intercept only. A statistically significant result (i.e., $p < .05$) means that the specified model predicts the dependent variable significantly better than the intercept-only model alone. As it turns out, the p-value is less than .01 in all 14 countries. For the sake of succinctness, these results will not be presented.

Measures of Predictive Power

For multinomial logistic regression, there are other measures of predictive power that are similar to R^2 in ordinary least-squares linear regression, hence the name Pseudo R^2. Like R^2 in multiple regression, Pseudo R^2 ranges between 0 and 1 with 1 suggesting the model accounts for 100% of variance in the outcome and 0 that it accounts for none of the variance. Table 40 shows different versions of Pseudo R^2 statistics that approximate the proportion of variance explained by the specified model rather than calculate it precisely (Cox &Snell, 1989; Nagelkerke, 1991;

McFadden, 1974). Among them, McFadden's measure has been recommended for multinomial and ordered logit (Allison, 2014).

Based on the McFadden Pseudo R^2, the specified model successfully explained more than 15% of variance in the outcome in the three Visegrád Group countries and the two East Asian countries – 18% in the Czech Republic, 17% in South Korea, 16% in the Slovak Republic and 15% in Poland and Japan. On the other hand, the specified model only explained less than 10% of variance in two Scandinavian countries and Ireland – 8% in Norway, 9% in Denmark and Ireland.

Table 40. Predictive Power of Multinomial Logistic Regression Models by Country

BEL	**Pseudo R-Square**
Cox and Snell	.152
Nagelkerke	.184
McFadden	.094

CZE	**Pseudo R-Square**
Cox and Snell	.269
Nagelkerke	.325
McFadden	.179

DNK	**Pseudo R-Square**
Cox and Snell	.132
Nagelkerke	.170
McFadden	.094

GEM	**Pseudo R-Square**
Cox and Snell	.193
Nagelkerke	.254
McFadden	.150

IRL	Pseudo R-Square
Cox and Snell	.151
Nagelkerke	.180
McFadden	.090

JPN	Pseudo R-Square
Cox and Snell	.202
Nagelkerke	.261
McFadden	.152

KOR	Pseudo R-Square
Cox and Snell	.236
Nagelkerke	.300
McFadden	.174

NLD	Pseudo R-Square
Cox and Snell	.206
Nagelkerke	.255
McFadden	.140

NOR	Pseudo R-Square
Cox and Snell	.143
Nagelkerke	.170
McFadden	.083

POL	Pseudo R-Square
Cox and Snell	.260
Nagelkerke	.301
McFadden	.150

SVK	Pseudo R-Square
Cox and Snell	.267
Nagelkerke	.310
McFadden	.157

SWE	Pseudo R-Square
Cox and Snell	.203
Nagelkerke	.253
McFadden	.140

UK	Pseudo R-Square
Cox and Snell	.159
Nagelkerke	.190
McFadden	.095

USA	Pseudo R-Square
Cox and Snell	.189
Nagelkerke	.240
McFadden	.135

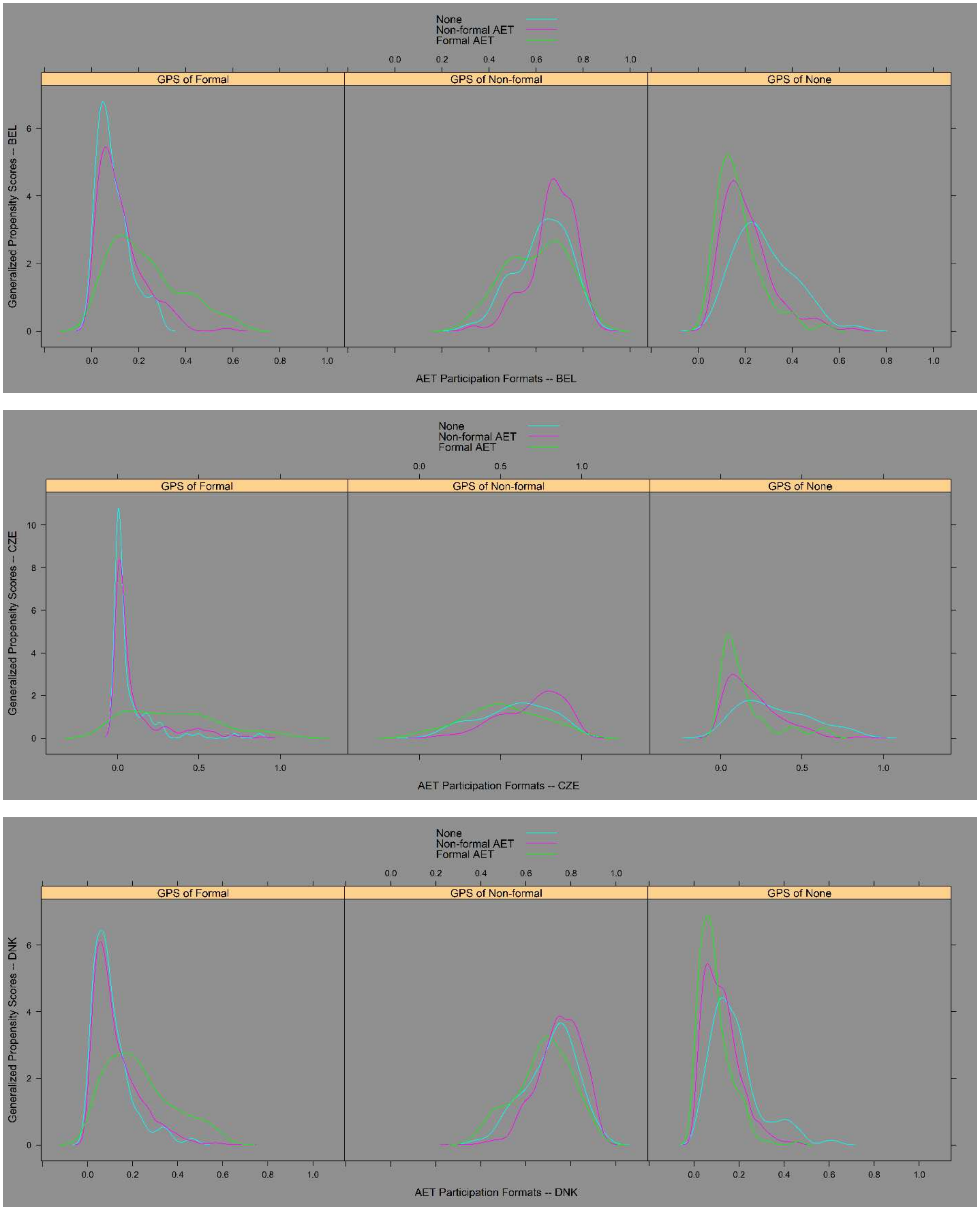
4.3.2 Assess Area of Common Support among the Different Formats of AET Participation

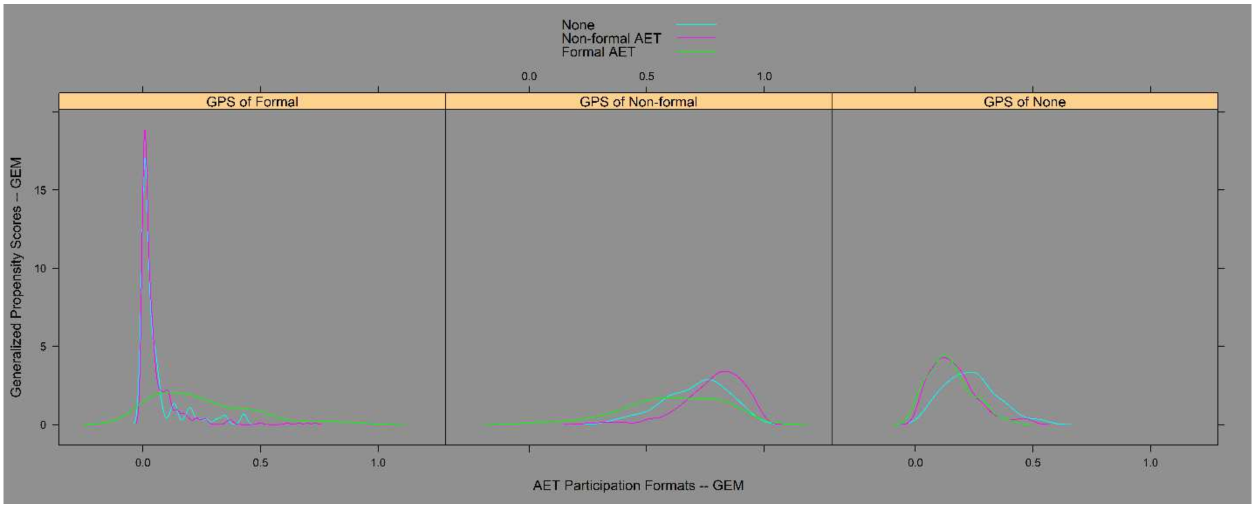
Once each individual in the sample is assigned a vector of three propensity scores (i.e., GPS) corresponding to the three formats of AET participation, it is important to assess the area of common support among the different formats of AET participation. As introduced in Chapter 2, for each vector of GPS and each pair of treatments being compared, the area of common support refers to the range of the GPS distribution where there are individuals with similar probabilities of receiving each treatment (Leite, 2017). Lack of common support may result in failure to obtain adequate covariate balance which, in turn, can lead to biased treatment effect estimates. For illustrative purpose, a kernel density plot is generated for each of the 14 selected countries that depicts the distribution of the estimated GPS for Non-formal, Formal and None AET participants respectively. Take Belgium for example, we first estimated the propensity scores for Formal AET participation for those who actually participated in None, Non-formal and Formal AET programs respectively, and then plot the distributions of GPS for Formal AET participation for each of the three types of AET participants. We repeated these steps for Non-formal and None AET participation. The top panel of Figure 36 presents the kernel density plots of the GPS distribution with scores for Formal AET participation on the left, Non-formal in the middle and None on the right.

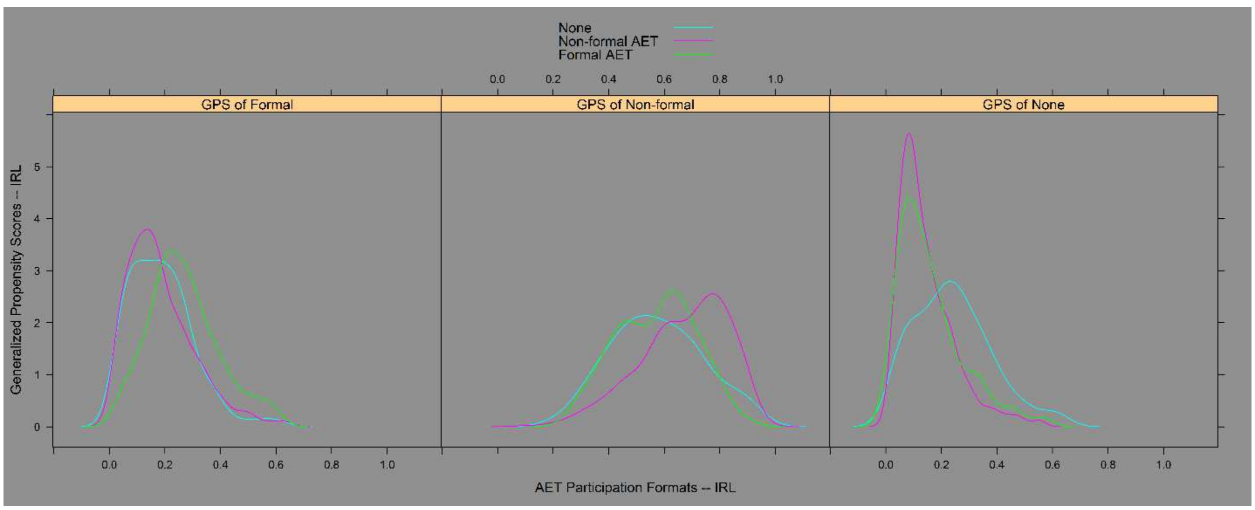
Although there is no explicit rule for “sufficient” area of common support, substantial overlap of the kernel density plots generally suggests that “the groups are sufficiently similar to support causal estimation of the treatment estimands” (McCaffrey et al, 2014). As shown on the left side of Figure 36, the distributions of estimated GPS for Formal AET participation appear to have adequate area of common support between actual Non-formal AET participants and actual None AET participants across the 14 selected countries. However, the overlap between actual Formal AET participants and other two groups seems to be weak in most countries, suggesting Formal AET participants may be distinct from Non-formal and None AET participants. Similar patterns were observed on the right side of Figure 36 where the distributions of estimated GPS for None AET participation were presented.

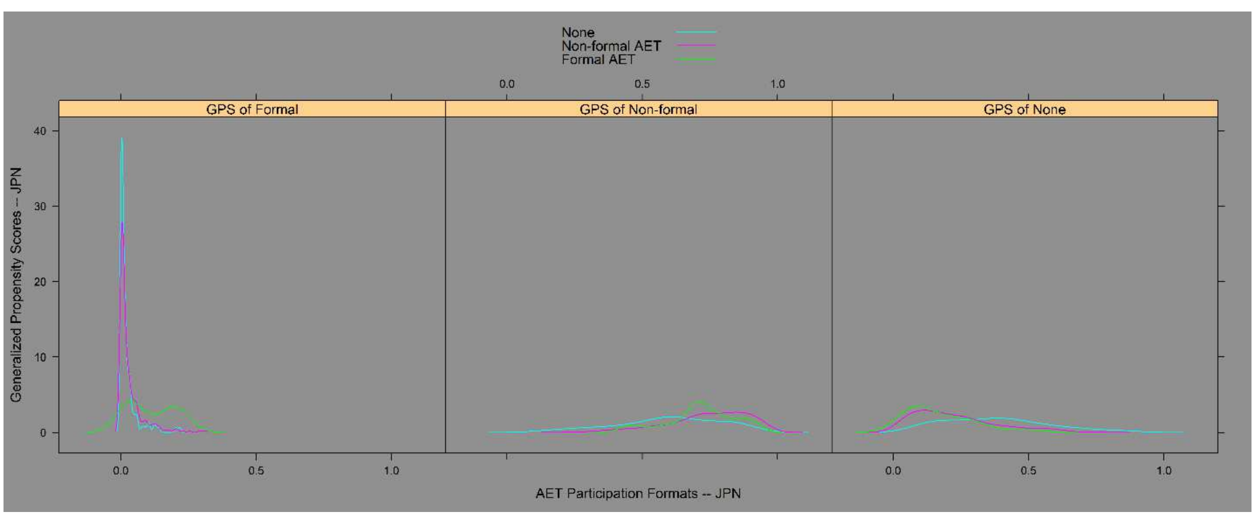
On the other hand, the distributions of estimated GPS for Non-formal AET participation have shown sufficient overlap among the three groups of AET participants (see the middle part of Figure 36). This means every individual in the sample could have participated in Non-formal AET programs. Such results are consistent with the RQ1 finding that Non-formal AET programs are the most popular among the sampled full-time professionals”. Further evidence of the adequacy of common support can be obtained later in the evaluation of covariate balance with propensity score weighting.

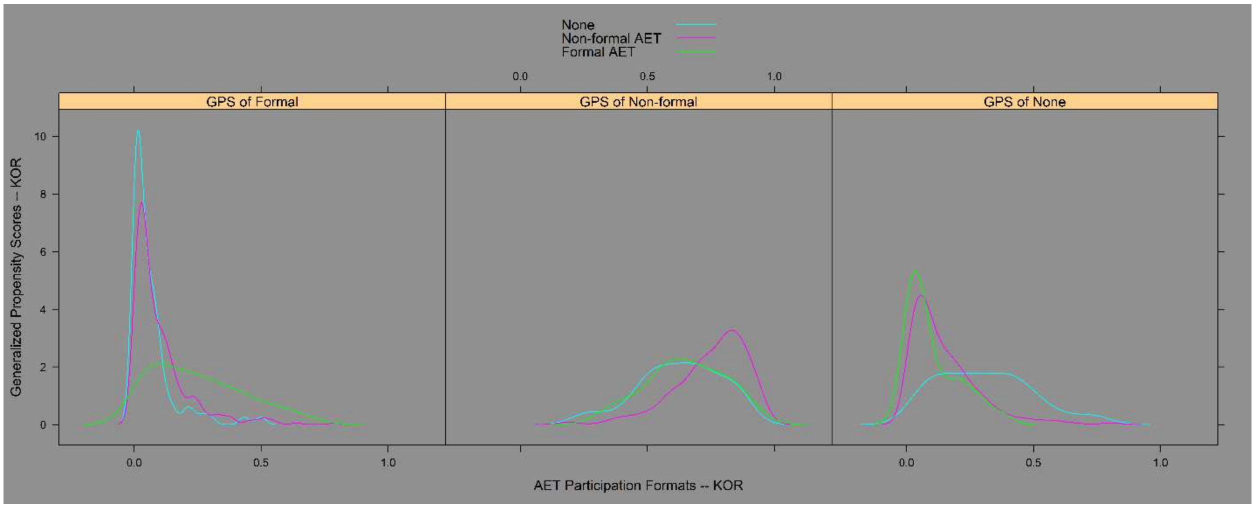
Figure 36. Kernel Density Plots of the Distribution of the Estimated GPS for Non-formal, Formal and None AET Participation

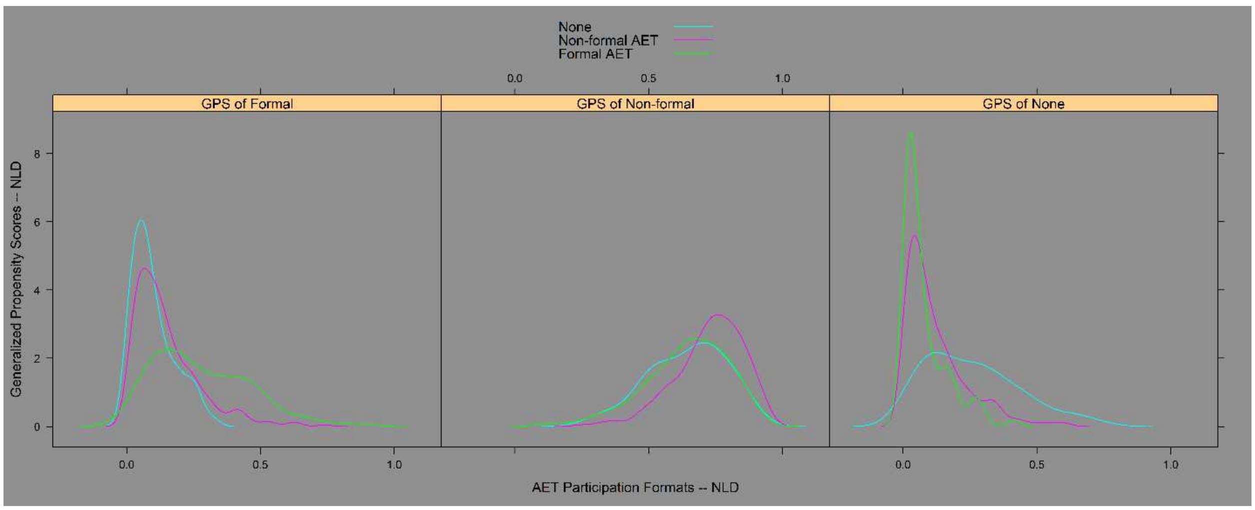


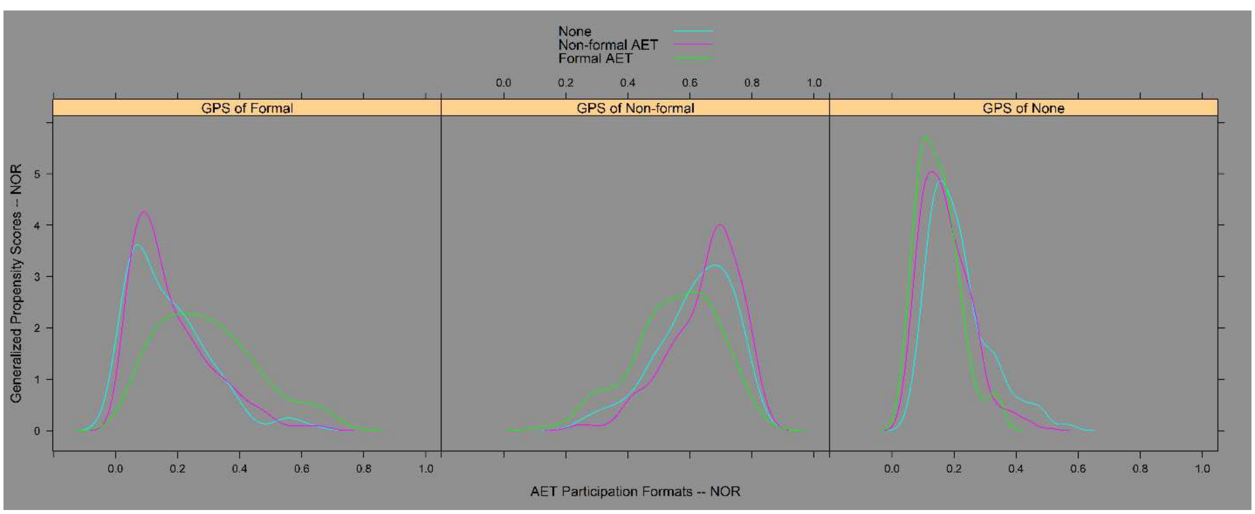

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- GEM
AET Participation Formats -- GEM

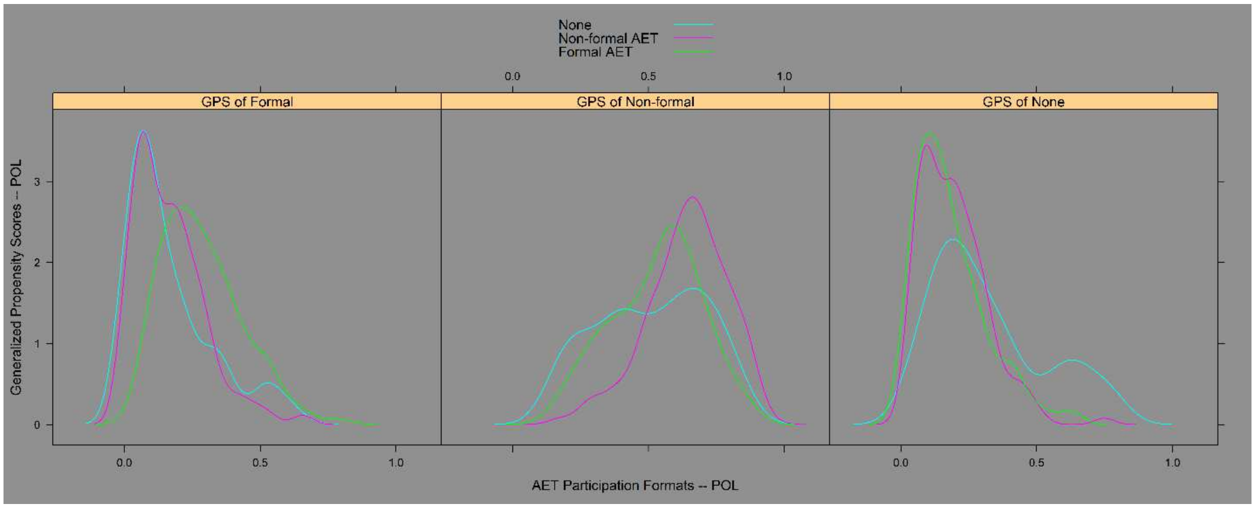

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- IRL
AET Participation Formats -- IRL


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- JPN
AET Participation Formats -- JPN

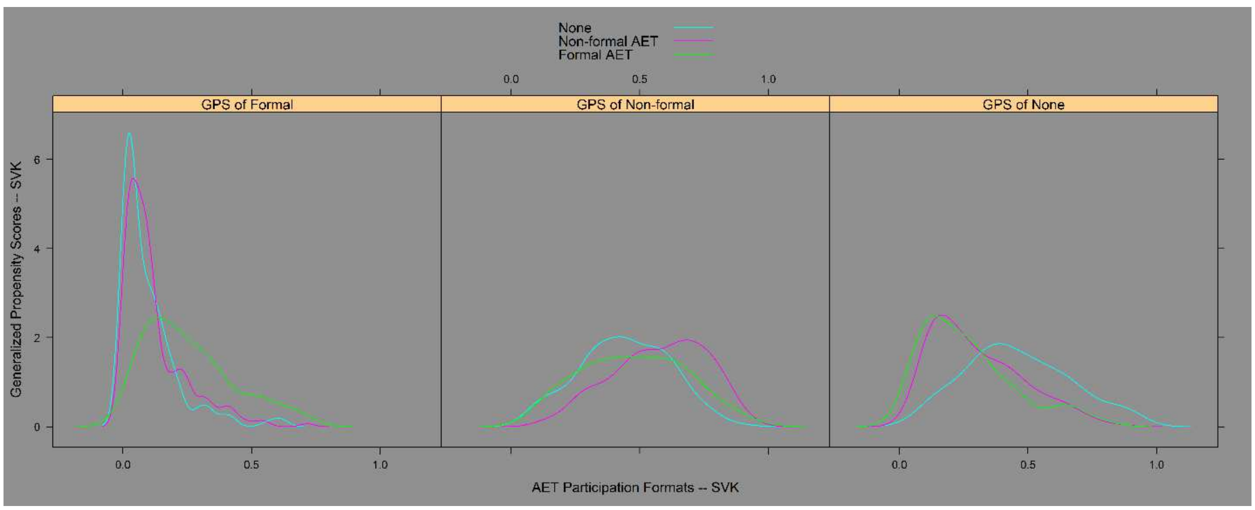

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- KOR
AET Participation Formats -- KOR
0.0
0.5
1.0
0
2
4
6
8
10


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- NLD
AET Participation Formats -- NLD
0.0
0.5
1.0
0
2
4
6
8

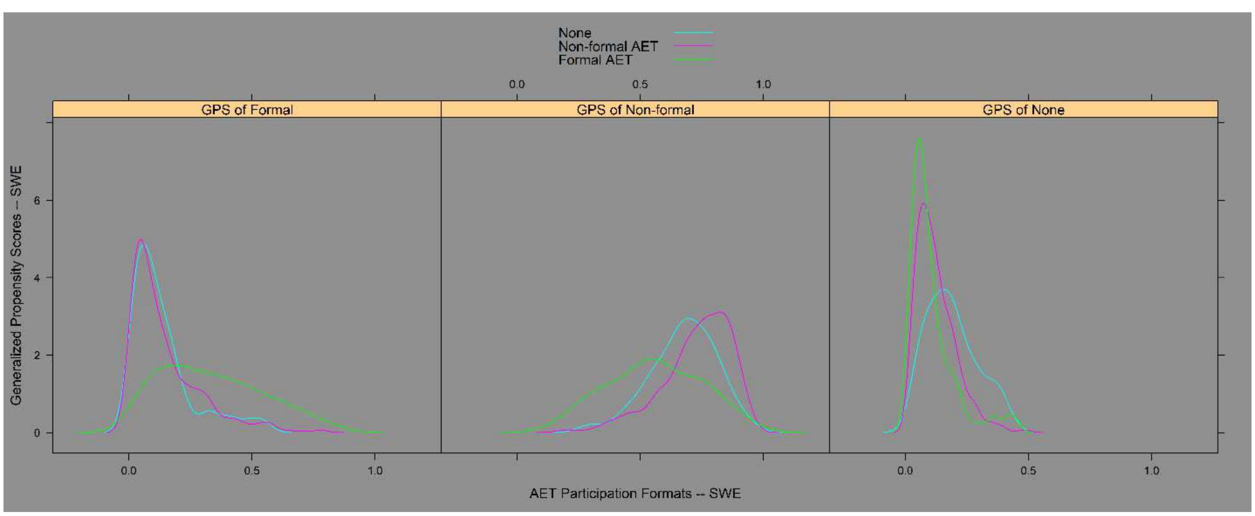

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- NOR
AET Participation Formats -- NOR
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5



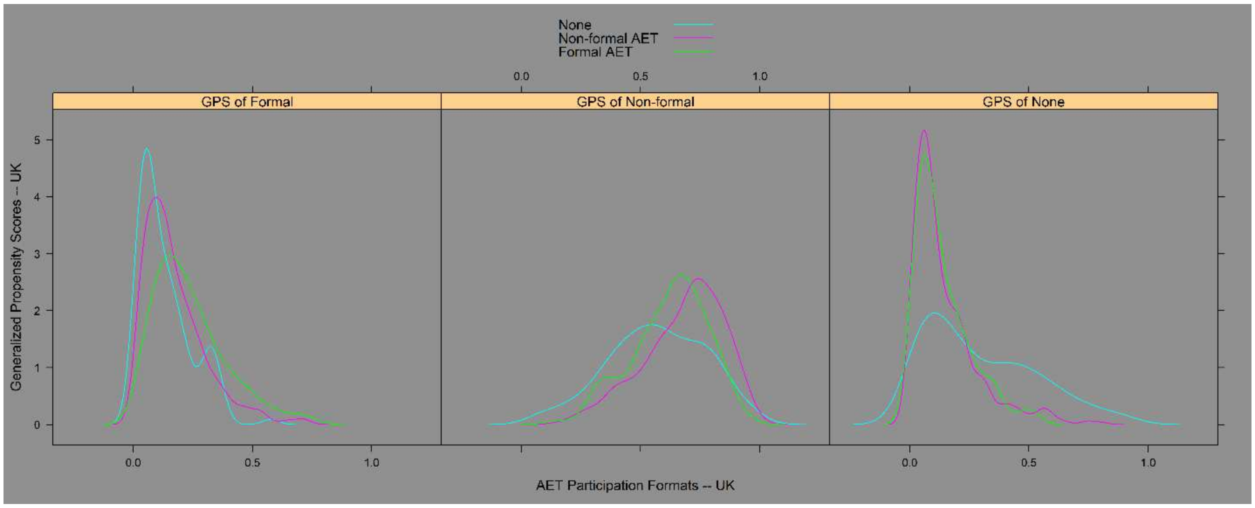
None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- POL
AET Participation Formats -- POL

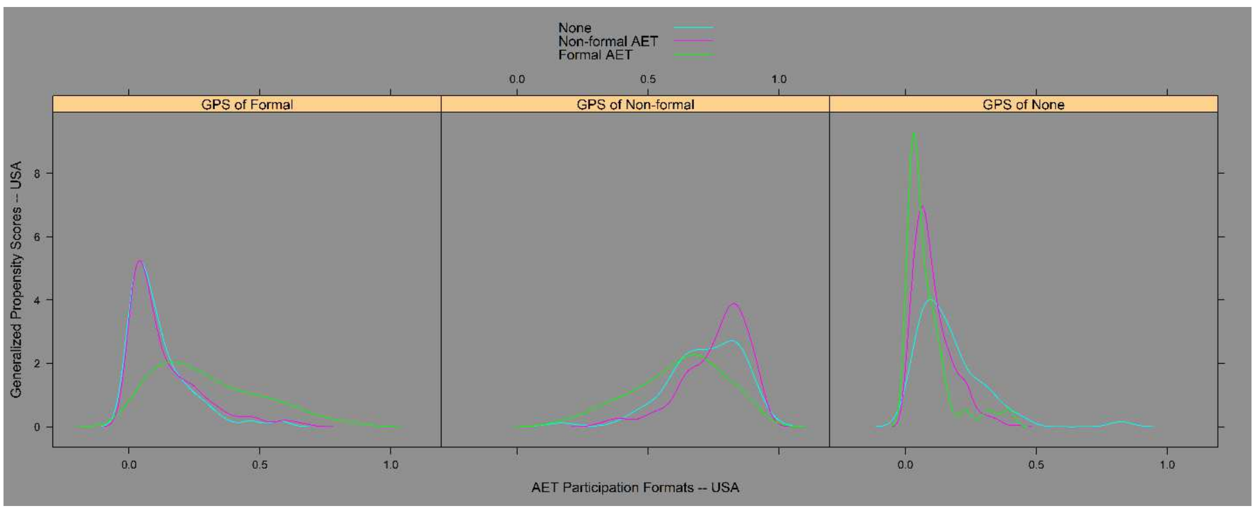


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- SVK
AET Participation Formats -- SVK



None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- SWE
AET Participation Formats -- SWE





4.3.3 Obtain Inverse Probability of Treatment Weights (IPTW) with GPS

To reduce selection bias in the estimation of the average treatment effects (ATE) with multiple treatment conditions, the inverse probability of treatment weighting (IPTW) method is used in this study. As discussed in Chapter 2, the weight assigned to an individual is defined as the inverse of the GPS for the condition that he/she was exposed to (Imbens, 2000). What is more, to account for the complex sampling design of PIAAC, we normalized the weights by first multiplying the IPTW by the final full sample weight from PIAAC (i.e., SPFWT0) and then splitting by the mean of the weights. As a result, the normalized weights should sum to the sample size in each country. Table 41 summarizes the normalized weights by full-time professionals' actual AET participation across the 14 selected countries.

Table 41. Summary of the Normalized Weights by Full-time Professionals' Actual AET Participation by Country

BEL	Min.	Median	Mean	Max.
None	.50	.92	1.07	4.19
Non-formal	.44	.74	.75	1.13
Formal	.60	1.70	2.08	6.51

* N = 514

CZE	Min.	Median	Mean	Max.
None	.05	.73	1.13	6.37
Non-formal	.02	.50	.96	7.12
Formal	.04	.71	.99	4.32

* N = 428

DNK	Min.	Median	Mean	Max.
None	.31	1.99	2.37	13.15
Non-formal	.06	.51	.55	1.75
Formal	.38	1.85	2.20	6.52

* N = 1238

GEM	Min.	Median	Mean	Max.
None	.86	1.92	2.18	4.31
Non-formal	.20	.47	.52	1.90
Formal	.54	2.53	3.06	10.39

* N = 513

IRL	Min.	Median	Mean	Max.
None	.22	.86	1.03	5.03
Non-formal	.08	.67	.71	2.46
Formal	.46	1.70	1.94	5.78

* N = 676

JPN	Min.	Median	Mean	Max.
None	.59	1.55	1.70	4.11
Non-formal	.38	.70	.74	1.66
Formal	.44	.81	.91	2.07

* N = 430

KOR	Min.	Median	Mean	Max.
None	.12	.93	1.18	4.74
Non-formal	.12	.55	.54	1.98
Formal	.82	3.58	3.61	10.95

* N = 508

NLD	Min.	Median	Mean	Max.
None	1.71	2.78	2.90	4.96
Non-formal	.20	.49	.52	1.11
Formal	.32	1.29	1.51	4.88

* N = 471

NOR	Min.	Median	Mean	Max.
None	.67	1.25	1.44	3.51
Non-formal	.49	.68	.69	1.67
Formal	.55	1.37	1.56	5.48

* N = 668

POL	Min.	Median	Mean	Max.
None	.06	.97	1.13	4.44
Non-formal	.06	.79	.77	2.60
Formal	.09	.71	1.42	9.56

* N = 513

SVK	Min.	Median	Mean	Max.
None	.23	.75	.86	3.90
Non-formal	.20	.67	.77	2.04
Formal	.50	1.82	2.20	10.64

* N = 435

SWE	**Min.**	**Median**	**Mean**	**Max.**
None	.82	2.83	2.85	6.38
Non-formal	.28	.62	.67	1.78
Formal	.37	.84	1.03	4.62

* N = 689

UK	**Min.**	**Median**	**Mean**	**Max.**
None	.03	1.12	1.88	10.57
Non-formal	.01	.51	.64	4.28
Formal	.02	.86	1.35	5.21

* N = 701

USA	**Min.**	**Median**	**Mean**	**Max.**
None	.33	.77	.89	5.30
Non-formal	.19	.59	.62	1.66
Formal	1.17	2.65	2.67	5.64

* N = 573

Because extreme weights could dramatically increase the variance of the treatment effect estimates (Leite, 2017), it is important to check the distribution of the weights. A rule of thumb for well-behaved weights would be a mean normalized weight close to 1 and a maximum normalized weight of < 10 (Cole & Hernan, 2008). Comparing the normalized weights against this rule of thumb suggests that the weights obtained are not extreme. Therefore, truncating the distribution of weights at the 99th or 95th percentile is not necessary for this study.

4.3.4 Evaluate Pairwise Covariate Balance Across Three Formats of AET Participation

As a crucial component of propensity score analysis, covariate balance assessment entails assessing "the degree to which conditioning on the propensity score has balanced measured baseline covariates between treatment groups" (Austin, 2019). According to Leite (2017), there are two approaches for the assessment of covariate balance with multiple treatment conditions:

(1) between each treatment condition and all the other conditions combined and (2) between all possible pairs of treatment conditions. Here, the latter (i.e., pairwise covariate balance) was adopted because it is important to identify precisely which pair of conditions does not have adequate covariate balance.

As mentioned in Chapter 2, there is no definitive guideline of what magnitude of standardized effect size is too large. In this study, we observe the general criterion that considered the standardized effect size of less than .20 as small, .40 as moderate and .60 as large (Cohen, 1988). Table 42 summarizes the absolute values of the standardized effect sizes between all possible pairs of AET participation formats for all 17 predictors across the 14 selected countries.

The first column of Table 42 refers to the standardized effect size between None and Non-formal AET participation, the second column refers to the standardized effect size between None and Formal AET participation and the third column refers to the standardized effect size between Non-formal and Formal AET participation. An examination of the last row of the first column shows that the maximum standardized effect size between None and Non-formal AET participation is well below .40 in all 14 countries. This indicates at least acceptable covariate balance between None and Non-formal AET participation for all 17 predictors across countries. Moreover, half of the 14 countries have achieved adequate covariate balance with the maximum standardized effect size standardized effect size less than .20 -- Belgium, Germany, Ireland, Norway, Denmark, the Slovak Republic and Poland (in ascending order).

Table 42. Summary of the Standardized Effect Size between Pairs of AET Participation Formats among Full-time Professionals by Country

BEL	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.03	.04	.12
Std.eff.sz.1	.00	.07	.08	.25
Std.eff.sz.2	.01	.08	.09	.25

CZE	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.08	.08	.26
Std.eff.sz.1	.00	.13	.13	.67
Std.eff.sz.2	.00	.14	.14	.50

DNK	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.04	.05	.19
Std.eff.sz.1	.00	.10	.10	.31
Std.eff.sz.2	.01	.05	.06	.22

GEM	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.05	.06	.16
Std.eff.sz.1	.00	.16	.18	.66
Std.eff.sz.2	.00	.14	.18	.71

IRL	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.05	.05	.16
Std.eff.sz.1	.00	.05	.06	.16
Std.eff.sz.2	.00	.04	.05	.14

JPN	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.05	.06	.21
Std.eff.sz.1	.00	.15	.21	.80
Std.eff.sz.2	.00	.14	.22	.86

KOR	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.10	.12	.29

Std.eff.sz.1	.01	.16	.21	.64
Std.eff.sz.2	.01	.12	.15	.56

NLD	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.06	.08	.35
Std.eff.sz.1	.01	.09	.10	.46
Std.eff.sz.2	.00	.05	.05	.18

NOR	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.03	.05	.16
Std.eff.sz.1	.00	.07	.09	.23
Std.eff.sz.2	.00	.05	.06	.14

POL	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.05	.06	.19
Std.eff.sz.1	.02	.14	.16	.37
Std.eff.sz.2	.00	.10	.20	.42

SVK	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.06	.06	.17
Std.eff.sz.1	.01	.15	.18	.53
Std.eff.sz.2	.02	.18	.19	.54

SWE	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.07	.07	.21
Std.eff.sz.1	.01	.07	.11	.34
Std.eff.sz.2	.00	.07	.09	.25

UK	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.07	.09	.36
Std.eff.sz.1	.00	.10	.13	.34
Std.eff.sz.2	.00	.08	.08	.22

USA	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.07	.07	.22

Std.eff.sz.1	.00	.13	.16	.44
Std.eff.sz.2	.01	.08	.12	.42

By contrast, the maximum standardized effect size between None and Formal AET participation is exceeding .40 in half of the 14 countries – the United States (.44), the Netherland (.46), the Slovak Republic (.53), South Korea (.64), Germany (.66), the Czech Republic (.67) and Japan (.80). However, an examination of the second-to-last row of the second column shows that the third quartile of the standardized effect size between None and Formal AET participation is well below .40 in all 14 countries. Therefore, the lack of covariate balance might not be very severe between None and Formal AET participation.

A scrutiny of the standardized effect sizes between None and Formal AET participation was conducted for each country and covariates did not achieve balance within .40 standard deviations are listed in Table 43 below.

Table 43. Countries with Maximum Standardized Effect Sizes > .4 between None and Formal AET Participation

Country	Covariates
The United States	GENDER
	EDU: less than B.A.
The Netherlands	ISCO2C: Health
The Slovak Republic	ISCO2C: Business & Administration
	AGE10BAND: 25-34
	GENDER
South Korea	AGE10BAND: 25-34
	PARED: at least one tertiary
	REARWORK
	READHOME
	PLANNING
Germany	**AGE10BAND: 25-34**
	TASKDISC
The Czech Republic	**EDU: B.A.**
Japan	PARED: at least one tertiary
	WRITEWORK
	READHOME
	WRITHOME
	NUMHOME
	ICTHOME

** Covariates that did not meet the .40 balance criteria for both None vs. Formal AET participation and Non-formal vs. Formal AET participation are bolded.*

Similarly, the maximum standardized effect size between Non-formal and Formal AET participation is exceeding .40 in half of the 14 countries – Poland (.42), the United States (.42), the Czech Republic (.50), the Slovak Republic (.54), South Korea (.56), Germany (.71), and Japan (.86). And the third quartile of the standardized effect size between Non-formal and Formal AET participation is well below .40 in all 14 countries, indicating a less severe lack of covariate balance. Table 44 below lists the covariates that did not achieve balance within .40 standard deviations for each country.

Table 44. Countries with Maximum Standardized Effect Sizes > .4 between Non-formal and Formal AET Participation

Country	Covariates
Poland	ISCO2C: Science & Engineering ISCO2C: Legal, Social & Cultural
The United States	**EDU: less than B.A.**
The Czech Republic	**EDU: B.A.**
The Slovak Republic	**AGE10BAND: 25-34**
South Korea	**READWORK** **READHOME**
Germany	**AGE10BAND: 25-34** AGE10BAND: 45-54
Japan	**WRITEWORK** **READHOME** **WRITHOME** **NUMHOME** **ICTHOME**

** Covariates that did not meet the .40 balance criteria for both None vs. Formal AET participation and Non-formal vs. Formal AET participation are bolded.*

In sum, across the three formats of AET participation, the balance criterion of .4 was not achieved with all covariates. According to McCaffrey et al. (2014), covariates that remain unbalanced after weighting could potentially confound estimated treatment effects. To address this challenge, Chapter 2 provides three remedial actions -- re-specifying the propensity score model, changing the propensity score estimation method, or including the covariates that did not reach the specified criteria for covariate balance in the outcome model. Because most of the remaining unbalances (i.e., differences between groups) are modest (i.e., not exceeding .60), it is advisable to include the unbalanced covariates in the outcome model (McCaffrey et al., 2014) so as to increase the robustness of the ATE estimators to misspecifications of the propensity score model.

4.3.5 Estimate Treatment Effects for Three Formats of AET Participation

When the imbalances remain after weighting, one common approach is to estimate treatment effects through weighted regression on treatment indicators and unbalanced covariates. It is a form of 'doubly robust' estimation because it yields consistent estimates of the treatment effects "if either the model for the outcome or the propensity score model is correct but not necessarily both' (Schafer & Kang, 2008). In other words, doubly robust estimation provides the best possible protection against possible errors in the propensity score model.

As described in Chapter 2, the ATE of K-1 treatments with respect to the reference condition can be estimated with a weighted regression model with K-1 dummy-coded treatment indicators (Leite, 2017):

$$Y_i = b_0 + \sum_{1}^{K-1} \boldsymbol{b}_k \, T_{ik} + e_i \qquad \textbf{Equation 14}$$

where $\mathbf{b_k}$ is the ATE of each treatment with respect to the reference condition and can be calculated as follows:

$$\boldsymbol{b}_k = \sum_{1}^{n} \boldsymbol{w_i y_i T_{ik}} \,/\, \sum_{1}^{n} \boldsymbol{w_i T_{ik}} \qquad \textbf{Equation 15}$$

To obtain the doubly robust estimation the ATE of K-1 treatments (with respect to the reference condition), covariates that did not reach the specified criteria for covariate balance were included as predictors in the weighted regression model, along with the K-1 dummy-coded treatment indicators. What is more, interactions between unbalanced covariates and the treatment condition are also included to capture any nonlinearities in their relationships to the different format of AET participation (McCaffrey et al., 2014).

For continuous outcomes, the outcome model can be expressed as:

$$Y_i = b_0 + \sum_1^{K-1} \boldsymbol{b}_k T_{ik} + \sum_1^{J} \boldsymbol{d_j}\, \boldsymbol{X_{ij}} + \sum_1^{K-1}\sum_1^{J} \gamma_{\boldsymbol{kj}}\, \boldsymbol{T_{ik}}\, \boldsymbol{X_{ij}} + e_i \qquad \textbf{Equation 16}$$

where $\boldsymbol{Y_i}$ is the continuous outcome, or the mean PS-TRE proficiency score for each individual. $\boldsymbol{b_k}$ stands for the estimated treatment effect for $\boldsymbol{T_{ik}}$ and K = 3 for this study. $\boldsymbol{d_j}$ represents the coefficient on unbalanced covariate $\boldsymbol{X_{ij}}$ and J varies for each country. $\gamma_{\boldsymbol{kj}}$ indicates the interaction between the treatment condition $\boldsymbol{T_{ik}}$ and the unbalanced covariate $\boldsymbol{X_{ij}}$. Lastly, $\boldsymbol{e_i}$ is the error term that accounts for the uncertainty in the model. Because the regression model can only estimate the effects of treatment compared to the reference category (i.e., None), the pairwise differences between treatment conditions are calculated. In this study, the estimate of the ATE of Non-formal AET participation relative to Formal AET participation is estimated by the difference between pairs of coefficients on the treatment indicators – the coefficient for the Non-formal AET participation indicator minus the coefficient for the Formal AET participation indicator.

For binary outcomes, the outcome model can be expressed as:

$$\text{Log}\left(\frac{\boldsymbol{P(Y_i=1)}}{\boldsymbol{P(Y_i=0)}}\right) = b_0 + \sum_1^{K-1} \boldsymbol{b}_k T_{ik} + \sum_1^{J} \boldsymbol{d_j}\, \boldsymbol{X_{ij}} + \sum_1^{K-1}\sum_1^{J} \gamma_{\boldsymbol{kj}}\, \boldsymbol{T_{ik}}\, \boldsymbol{X_{ij}} \qquad \textbf{Equation 16}$$

where $\boldsymbol{Y_i}$ is the binary outcome, or whether the mean PS-TRE proficiency score for each individual is at or above the 75th percentile of the PS-TRE score distribution (i.e., top quartile) in the corresponding country. The model estimates the effects of treatment conditions and unbalanced covariates on the logit of scoring in the top quartile of the PS-TRE score distribution.

Table 45 below lists the relationships between the different formats of AET participation and full-time professionals' PS-TRE proficiency scores for each of the 14 selected countries, after controlling for unbalanced covariates (and interactions between unbalanced covariates and the treatment condition). As it turns out, AET participation (either Formal or Non-formal) is not significantly associated with full-time professionals' PS-TRE proficiency scores in all but one country – Denmark. In Denmark, full-time professionals who participated in Non-formal AET programs in the 12 months preceding the survey scored 27.33 points higher than their counterparts who participated in None ($p < .01$), controlling for propensity scores for Non-formal AET participation, unbalanced covariates and interactions between unbalanced covariates and the treatment condition. Similarly, full-time professionals who participated in Formal AET programs in the 12 months preceding the survey scored 26.63 points higher than their counterparts who participated in None ($p = .01$), controlling for propensity scores for Formal AET participation, unbalanced covariates and interactions between unbalanced covariates and the treatment condition. There is no significant difference in PS-TRE proficiency scores between Non-formal and Formal AET participants.

Table 45. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' PS-TRE Scores

BEL	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	20.41	22.02	.93	.35
Formal vs. None	37.61	24.71	1.56	.12
Non-formal vs. Formal	-17.20	24.41	-.70	.48

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

CZE	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	6.38	9.61	.66	.51
Formal vs. None	1.57	13.64	.12	.91
Non-formal vs. Formal	4.81	12.67	.38	.70

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

DNK	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	27.33	9.55	2.86	.004**
Formal vs. None	26.63	9.94	2.68	.01**
Non-formal vs. Formal	.71	5.10	.14	.89

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

GEM	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	-17.12	11.90	-1.44	.15
Formal vs. None	-24.65	22.18	-1.11	.27
Non-formal vs. Formal	7.54	20.39	.37	.71

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

IRL	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	-7.54	7.64	-.99	.32
Formal vs. None	-7.60	8.16	-.93	.35
Non-formal vs. Formal	.06	8.95	.01	.99

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-8.07	21.55	-.37	.71
Formal vs. None	-33.59	58.79	-.57	.57
Non-formal vs. Formal	25.52	59.44	.43	.67

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

KOR	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-30.03	21.74	-1.38	.17
Formal vs. None	22.17	27.21	.81	.42
Non-formal vs. Formal	-52.19	24.78	-2.11	.04

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

NLD	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-6.81	13.55	-.46	.65
Formal vs. None	12.30	17.57	.70	.48
Non-formal vs. Formal	-18.48	12.14	-1.52	.13

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

NOR	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-3.54	6.01	-.59	.56
Formal vs. None	15.54	9.36	1.66	.10
Non-formal vs. Formal	-19.08	9.41	-2.03	.04*

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

POL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-20.77	32.73	-.63	.53
Formal vs. None	22.72	22.19	1.02	.31
Non-formal vs. Formal	-43.49	33.72	-1.29	.20

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SVK	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	3.37	9 .90	.34	.73
Formal vs. None	15.98	19.04	.84	.40
Non-formal vs. Formal	-12.61	18.40	-.69	.49

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SWE	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	16.55	20.31	.81	.42
Formal vs. None	15.43	26.42	.58	.56
Non-formal vs. Formal	1.13	22.01	.05	.96

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

UK	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	7.94	19.11	.42	.68
Formal vs. None	7.35	15.81	.46	.64
Non-formal vs. Formal	.59	20.09	.03	.98

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

USA	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	28.18	29.02	.97	.33
Formal vs. None	3.46	33.63	.10	.92
Non-formal vs. Formal	24.72	25.50	.97	.33

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

Turning to the binary outcome (i.e., whether the mean PS-TRE proficiency score for each individual is in the top quartile of the PS-TRE score distribution in the corresponding country), logistic regression models yield highly consistent results. For the sake of brevity, only countries with significant results are presented in Table 46. At the .01 level, there seem to be some significant results in Japan and Poland with the binary outcome. However, these results from the logistic regression models ought to be treated with extreme caution due to the relative few Formal AET participants. For example, in Japan, there are only 19 Formal participants, compared to 297 Non-formal participants and 114 non-participants. Please refer to Figure 33 for more information on the distribution of the three formats of AET participation across countries.

At the .05 level, however, Denmark stands out as the only country where both formats of AET participation (vs. None) are significantly and positively associated full-time professionals' probability of scoring in the top quartile of the PS-TRE score distribution (p = .02). Specifically, after partialling out the variability due to the other variables in the model, the odds of scoring in the top quartile of the PS-TRE score distribution for Non-formal AET participants are 437% higher than those who participated in None – exp(1.68) = 5.37. And the odds of scoring in the top quartile of the PS-TRE score distribution for Formal AET participants are 442% higher than those who participated in None – exp(1.69) = 5.42. There is no significant difference between Non-formal and Formal AET participants in terms of the odds of scoring in the top quartile of the PS-TRE score distribution (p = .98).

Table 46. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the PS-TRE Score Distribution

DNK	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	1.68	.71	2.37	.02*
Formal vs. None	1.69	.73	2.31	.02*
Non-formal vs. Formal	-.01	.33	-.03	.98

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-2.60	3.71	-.70	.48
Formal vs. None	-33.07	11.99	-2.76	.01**
Non-formal vs. Formal	30.47	11.86	2.57	.01**

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

POL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-18.47	6.09	-3.03	.003**
Formal vs. None	2.80	1.83	1.53	.13
Non-formal vs. Formal	-21.26	6.00	-3.54	.000***

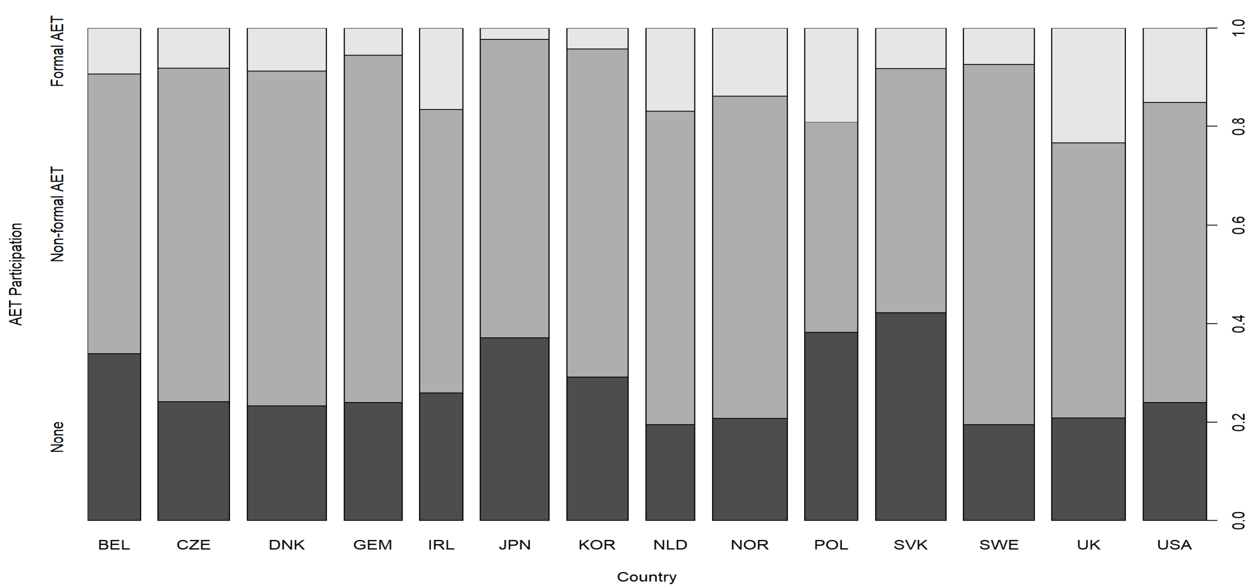
*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

4.4 Research Question Four

4.4.1 Parallel Analyses with Full-time Associates from the 14 Selected Countries

Corresponding to RQ3, which starts from predicting full-time professionals' probabilities of participating in the different formats of AET programs, the first part of research question four runs the same set of multinomial logistic regression models to predict full-time associates' probabilities of participating in the different formats of AET programs. First, Figure 37 (or Figure 33b) shows the distributions of full-time associates' participation in the three formats of AET programs by country. As is the case with full-time professionals, the 14 selected countries differ substantially not only in the distribution of participation rates across the three formats of AET programs, but also in the size of the sampled full-time associates (N = 6,146). It is worth noting that participation rates of Non-formal AET programs are highest in both samples.

Figure 37. Overview of Full-time Associates' Participation in the Different Formats of AET Programs by Country (Figure 33b, corresponding to Figure 33)



The remainder of the analysis for RQ4 replicates the steps used in RQ3 to estimate treatment effects for the three formats of AET participation. First, predicted probabilities for the three formats of AET participation are summarized by full-time associates' actual AET participation for each of the 14 selected countries. As shown in Table 47 (or Table 38b), each row represents the actual participants of one format of AET programs while each column represents the estimated probability for one format of AET participation. Across countries, the estimated probabilities of Non-formal AET participation are the highest across all three types of AET participants, which is consistent with the participation patterns observed with full-time professionals. Germany remains one of the countries with the highest propensity scores for Non-formal AET participation among actual Non-formal AET participants (.72, only second to .73 in South Korea) while the Slovak Republic remains to have the lowest (.53).

Table 47. Summary of the Estimated GPS Matrix for Full-time Associates by Their Actual AET Participation by Country (Table 38b, corresponding to Table 38)

BEL	None	Non-formal	Formal
None Participants	**.46**	.45	.08
Non-formal Participants	.27	**.66**	.08
Formal Participants	.31	.49	**.20**

CZE	None	Non-formal	Formal
None Participants	**.34**	.62	.05
Non-formal Participants	.24	**.72**	.04
Formal Participants	.19	.57	**.23**

DNK	None	Non-formal	Formal
None Participants	**.31**	.63	.06
Non-formal Participants	.21	**.70**	.08
Formal Participants	.15	.61	**.24**

GEM	None	Non-formal	Formal
None Participants	**.35**	.62	.03
Non-formal Participants	.23	**.72**	.05
Formal Participants	.15	.60	**.25**

IRL	None	Non-formal	Formal
None Participants	**.38**	.51	.11
Non-formal Participants	.21	**.65**	.15
Formal Participants	.17	.61	**.22**

JPN	None	Non-formal	Formal
None Participants	**.46**	.53	.02
Non-formal Participants	.31	**.67**	.02
Formal Participants	.28	.59	**.13**

KOR	None	Non-formal	Formal
None Participants	**.45**	.52	.04
Non-formal Participants	.23	**.73**	.04
Formal Participants	.25	.58	**.17**

NLD	None	Non-formal	Formal
None Participants	**.28**	.61	.11
Non-formal Participants	.19	**.67**	.14
Formal Participants	.13	.54	**.33**

NOR	None	Non-formal	Formal
None Participants	**.31**	.59	.10
Non-formal Participants	.20	**.67**	.13
Formal Participants	.16	.61	**.24**

POL	None	Non-formal	Formal
None Participants	**.48**	.44	.08
Non-formal Participants	.33	**.56**	.11
Formal Participants	.27	.42	**.31**

SVK	None	Non-formal	Formal
None Participants	**.46**	.47	.07
Non-formal Participants	.38	**.53**	.08
Formal Participants	.32	.47	**.21**

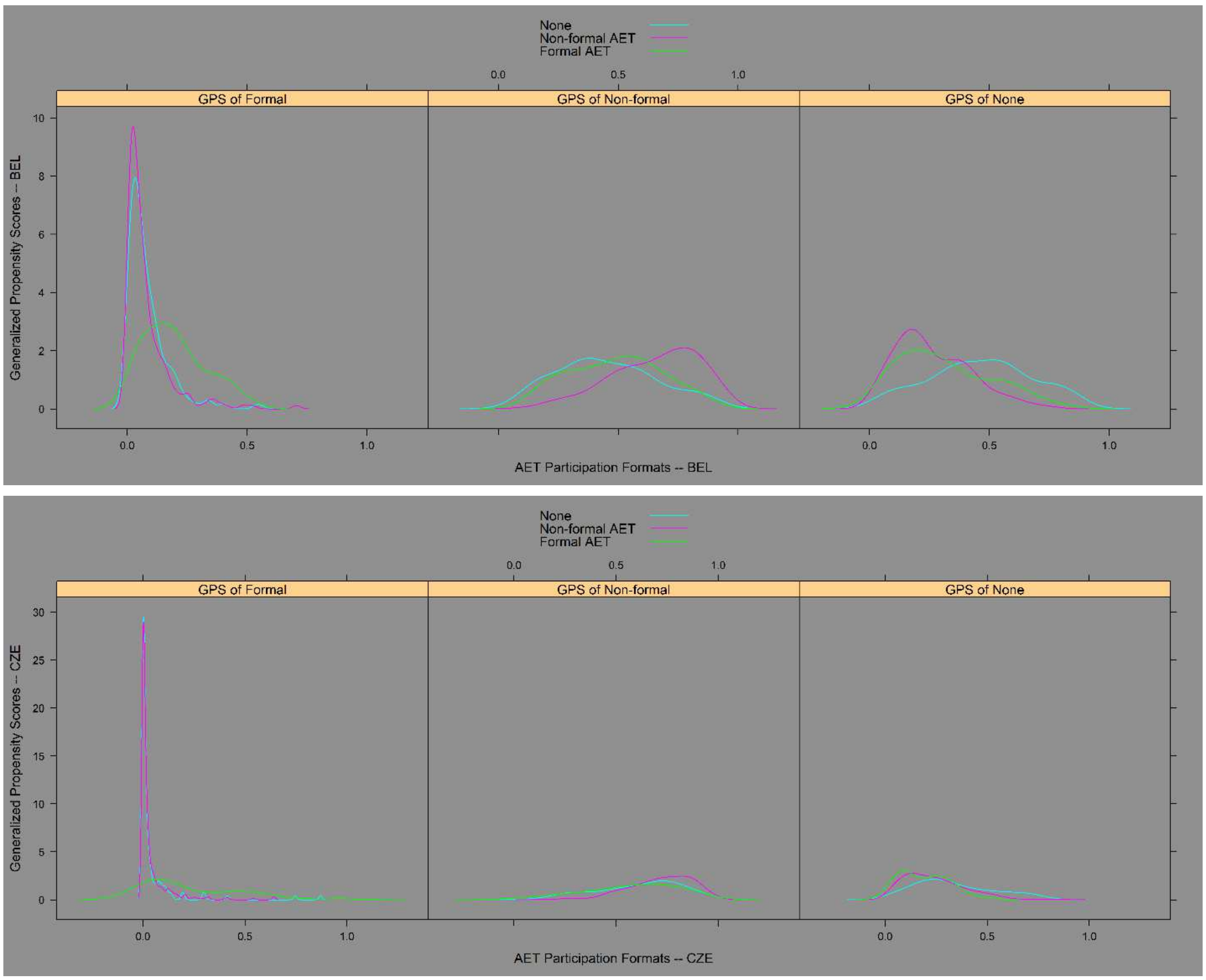
SWE	None	Non-formal	Formal
None Participants	**.26**	.69	.05
Non-formal Participants	.19	**.75**	.06
Formal Participants	.14	.59	**.27**

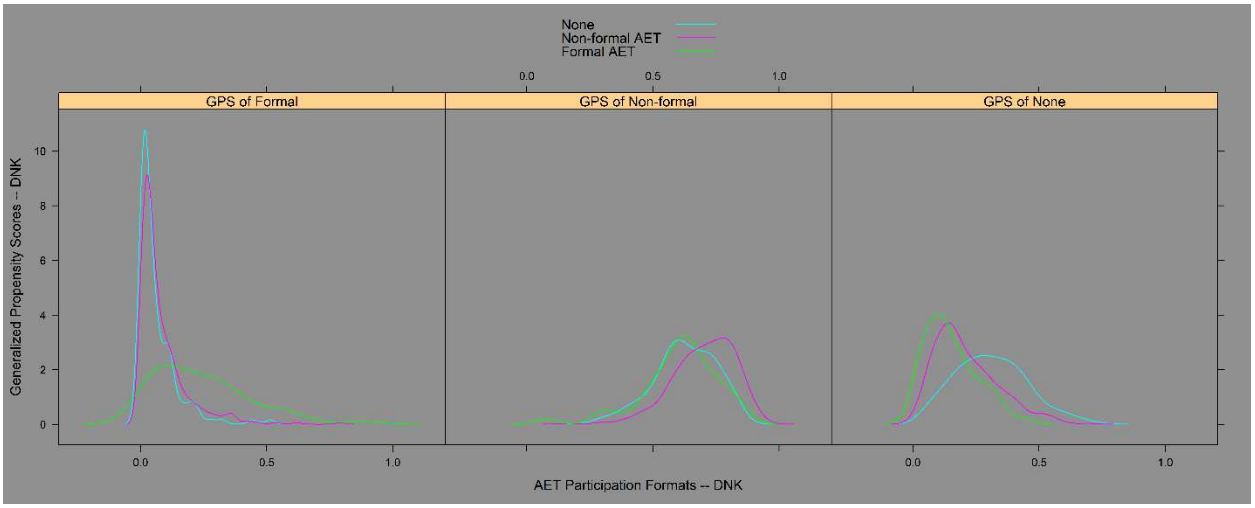
UK	None	Non-formal	Formal
None Participants	**.34**	.45	.21
Non-formal Participants	.20	**.55**	.25
Formal Participants	.18	.48	**.34**

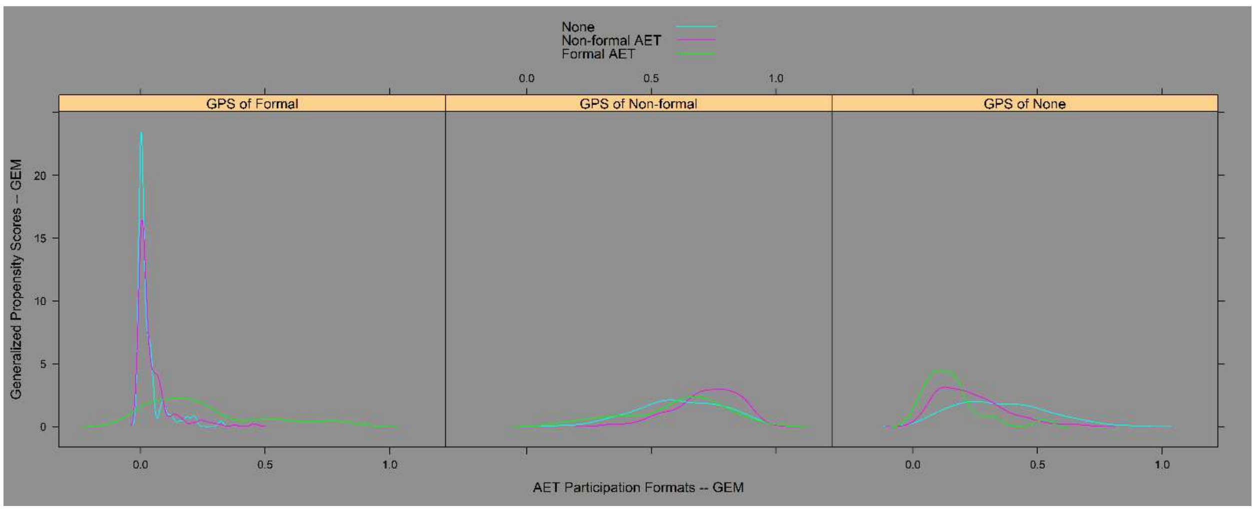
USA	None	Non-formal	Formal
None Participants	**.36**	.52	.12
Non-formal Participants	.23	**.65**	.12
Formal Participants	.20	.52	**.28**

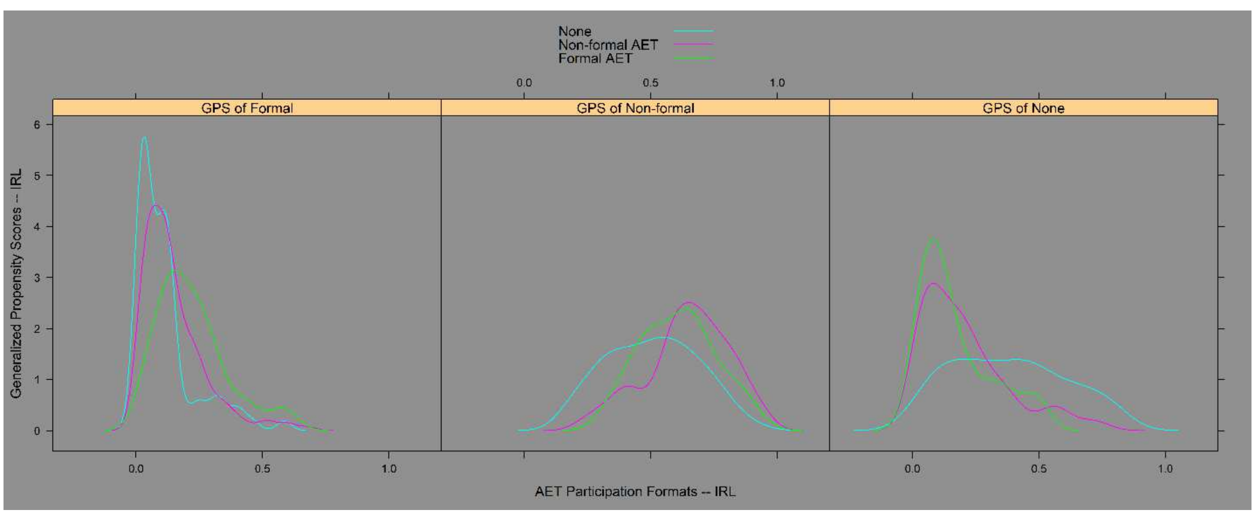
Once a vector of three propensity scores (i.e., GPS) is assigned to each individual, we assess the area of common support among the different formats of AET participation. Again, kernel density plots are deployed to demonstrate the distribution of the estimated GPS for Non-formal, Formal and None AET participants respectively. As is the case with full-time professionals, the distributions of estimated GPS for all three formats of AET participation overlap well between actual Non-formal AET participants and actual None AET participants across the 14 selected countries. However, the lack of overlap between actual Formal AET participants and other two groups observed for full-time professionals is observed for the sample of full-time associates, indicating that Formal AET participants are distinct from other participants in both samples. Please refer to Figure 38 (or Figure 36b) for a complete demonstration of the kernel density plot for each of the 14 selected countries.

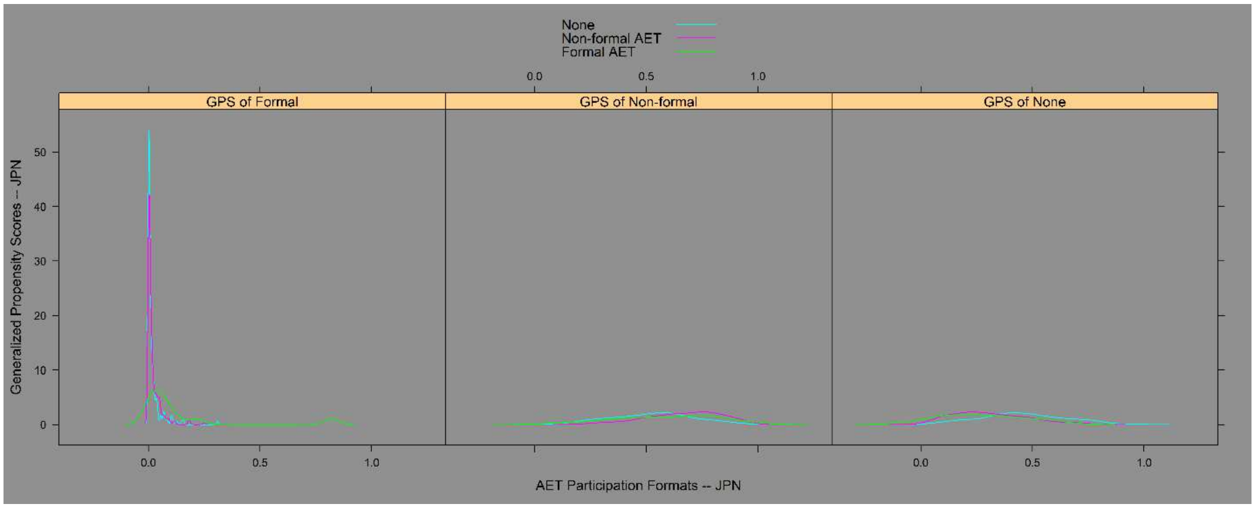
Figure 38. Kernel Density Plots of the Distribution of the Estimated GPS for Non-formal, Formal and None AET Participation (Figure 36b, corresponding to Figure 36)

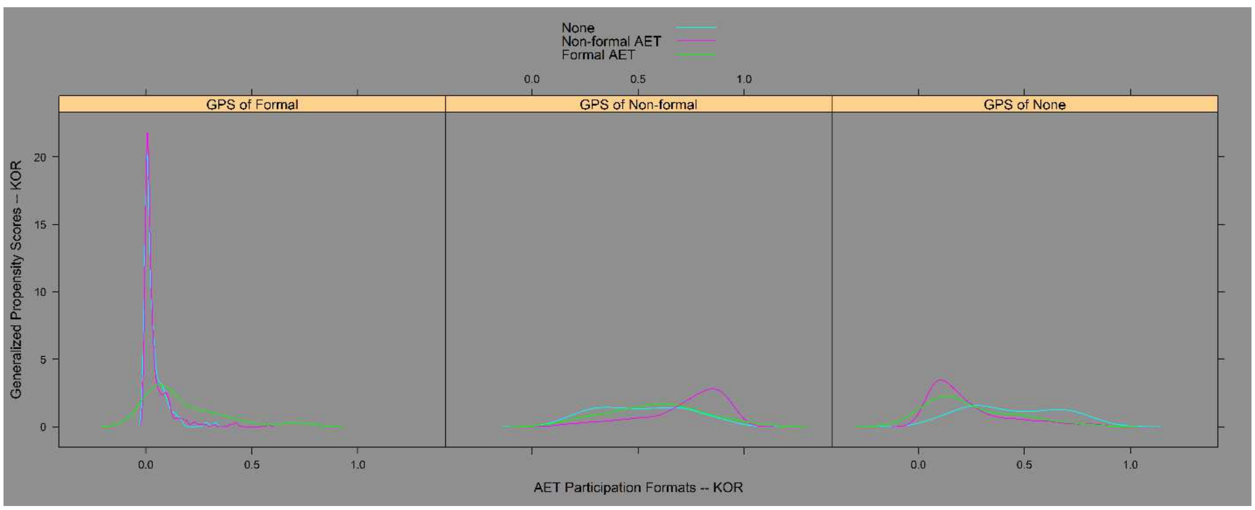


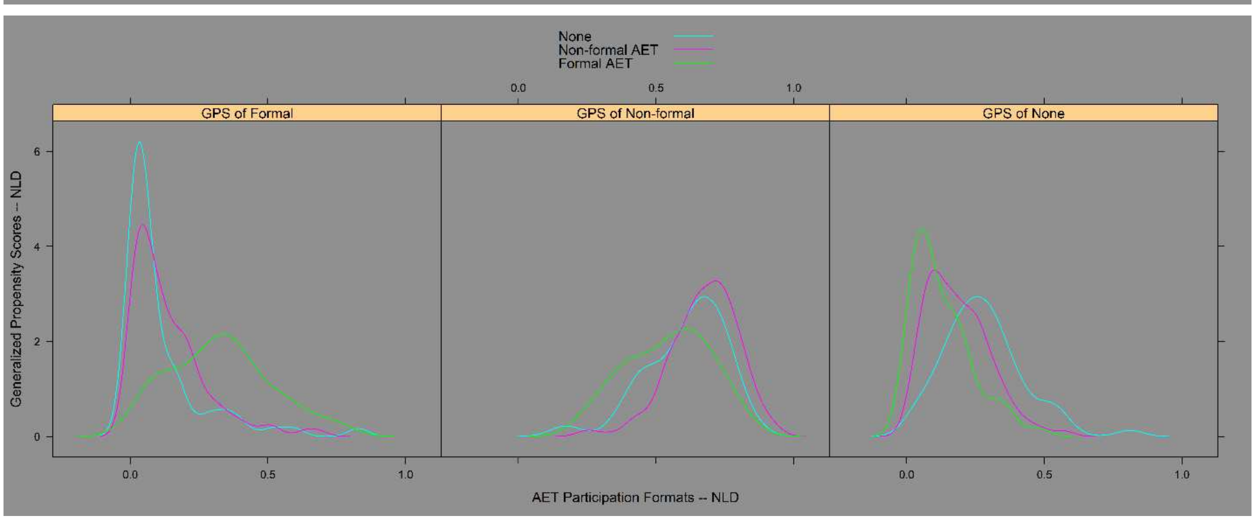

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- DNK
AET Participation Formats -- DNK

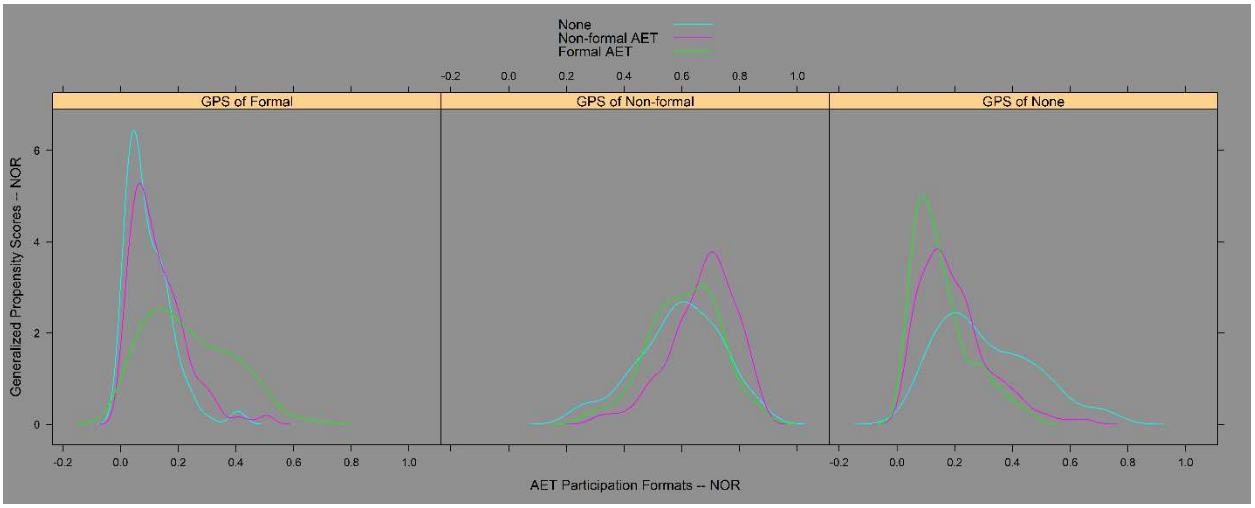

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- GEM
AET Participation Formats -- GEM

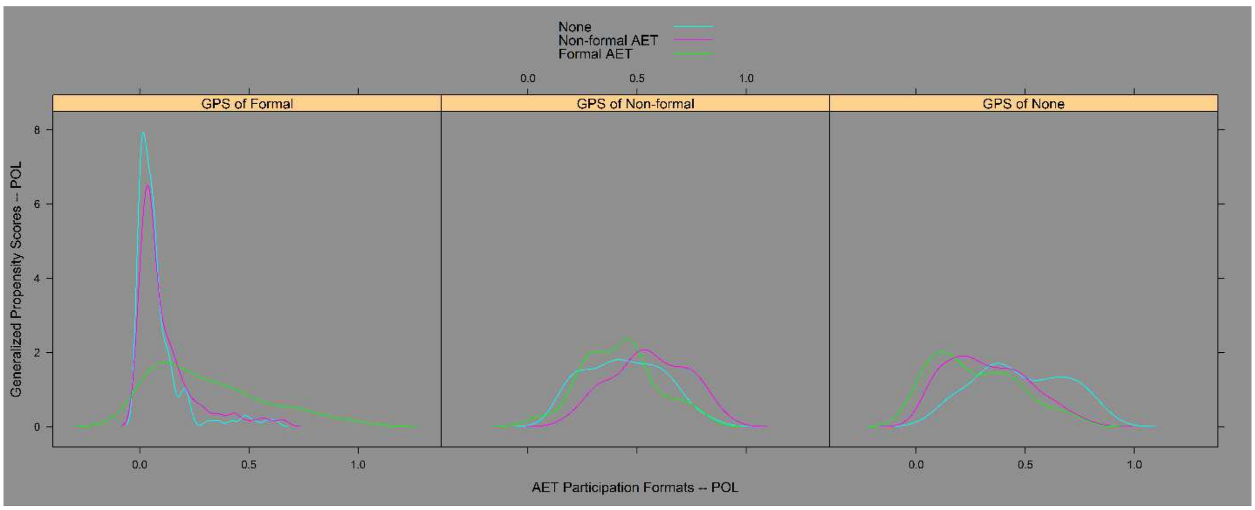

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- IRL
AET Participation Formats -- IRL

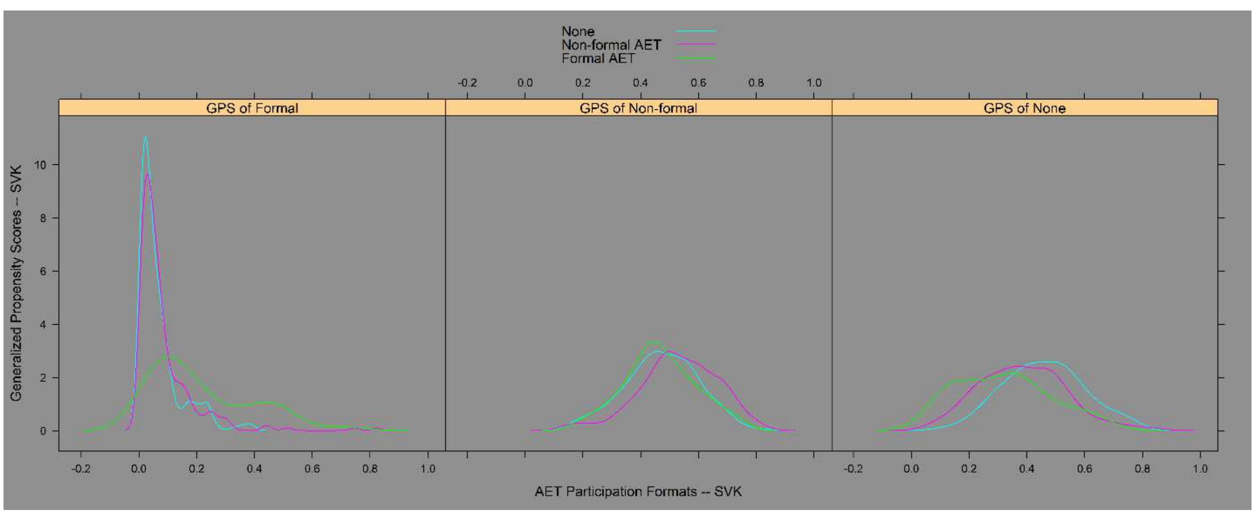

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- JPN
AET Participation Formats -- JPN

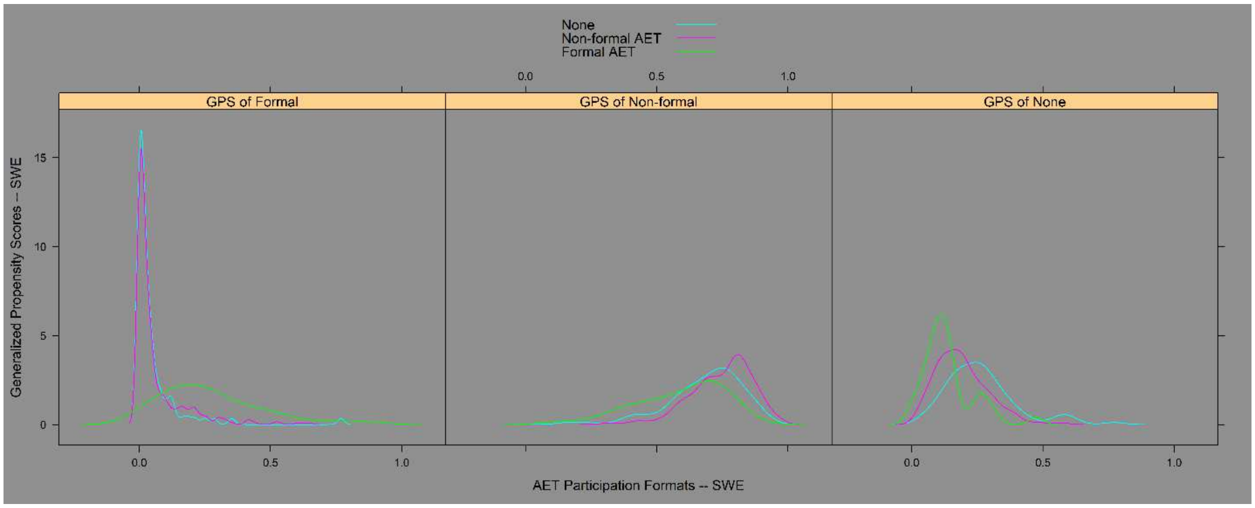

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- KOR
AET Participation Formats -- KOR

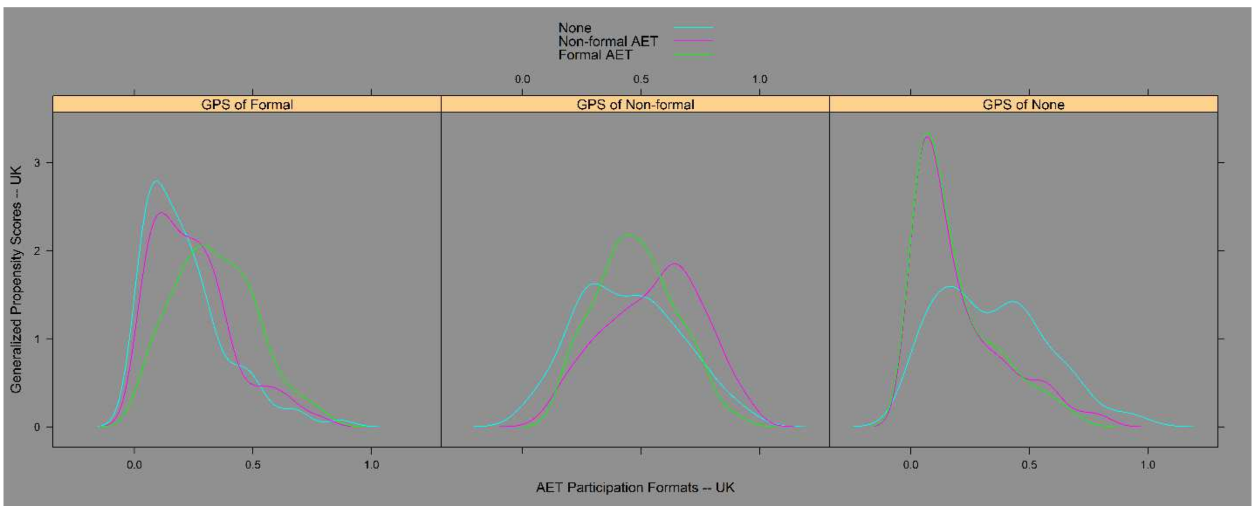

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- NLD
AET Participation Formats -- NLD

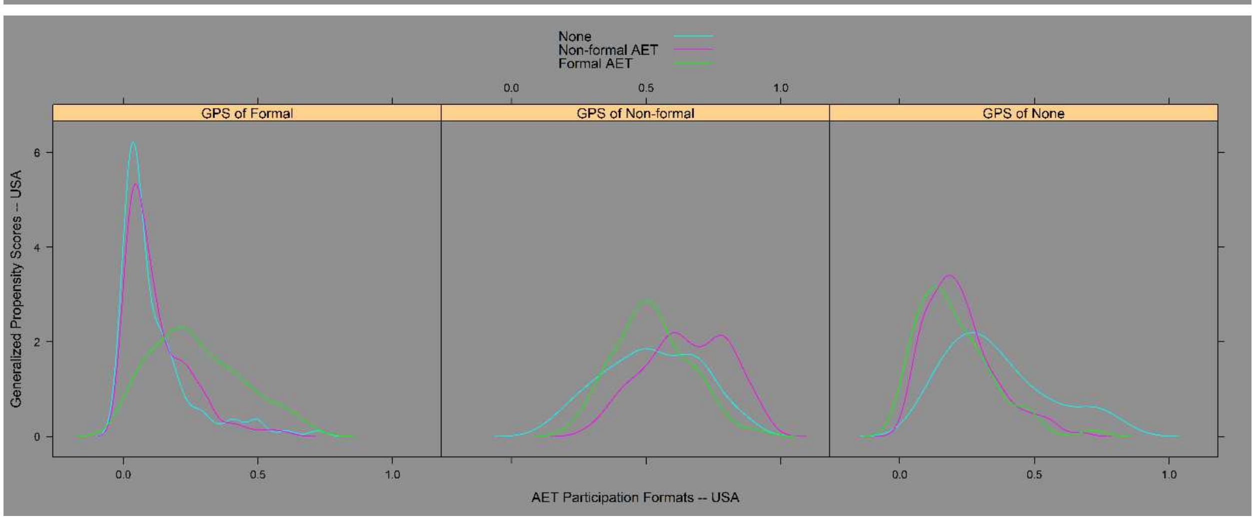

None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- NOR
AET Participation Formats -- NOR


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- POL
AET Participation Formats -- POL


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- SVK
AET Participation Formats -- SVK


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- SWE
AET Participation Formats -- SWE


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- UK
AET Participation Formats -- UK


None
Non-formal AET
Formal AET
GPS of Formal
GPS of Non-formal
GPS of None
Generalized Propensity Scores -- USA
AET Participation Formats -- USA

Just as with RQ3, we then obtain the normalized weights via splitting the product of IPTW and SPFWT0 (i.e., final full sample weight from PIAAC) by the mean of the weights. Please see Table 48 (or Table 41b) for summaries of the normalized weights by full-time associates' actual AET participation across the 14 selected countries. A check of the magnitudes of the weights against the rule of thumb that well-behaved weights should be a mean normalized weight close to 1 and a maximum normalized weight of < 10 confirms that the normalized weights for Non-formal and None AET participation are within a reasonable range in all 14 countries, while the normalized weights for Formal AET participation exceed the threshold of 10 in two countries – South Korea and the Slovak Republic (Cole & Hernan, 2008). Such results are not unexpected given the relative few Formal AET participants in these countries. Please refer to Figure 37 (or Figure 33b) for more information on the distribution of the three formats of AET participation across countries. What is more, the third quartile of normalized weights for Formal AET participation are well below 10 in all 14 countries. Therefore, we probably do not need to worry too much about increased variance of the treatment effect estimates due to extreme weights.

Table 48. Summary of the Normalized Weights by Full-time Associates' Actual AET Participation by Country (Table 41b, corresponding to Table 41)

BEL	Min.	Median	Mean	Max.
None	.40	.82	.92	2.79
Non-formal	.27	.50	.51	.89
Formal	2.77	4.32	4.29	5.67

* N = 352

CZE	Min.	Median	Mean	Max.
None	.08	1.13	1.39	5.87
Non-formal	.03	.51	.80	4.71
Formal	.09	.93	1.50	8.22

* N = 482

DNK	**Min.**	**Median**	**Mean**	**Max.**
None	.21	.99	1.02	2.48
Non-formal	.17	.84	.87	2.11
Formal	.51	1.62	1.97	6.02

* N = 546

GEM	**Min.**	**Median**	**Mean**	**Max.**
None	.32	1.45	1.65	4.99
Non-formal	.15	.74	.80	2.54
Formal	.43	.62	.69	1.33

* N = 376

IRL	**Min.**	**Median**	**Mean**	**Max.**
None	.26	.88	.89	2.19
Non-formal	.18	.88	.99	3.32
Formal	.25	1.12	1.21	3.70

* N = 290

JPN	**Min.**	**Median**	**Mean**	**Max.**
None	.44	1.08	1.25	3.66
Non-formal	.44	.80	.85	1.87
Formal	.40	.92	1.01	2.81

* N = 449

KOR	**Min.**	**Median**	**Mean**	**Max.**
None	.19	.83	1.03	3.91
Non-formal	.09	.48	.52	1.63
Formal	3.63	8.39	8.36	15.31

* N = 422

NLD	Min.	Median	Mean	Max.
None	1.03	2.05	2.51	3.25
Non-formal	.31	.54	.57	1.29
Formal	.62	1.25	1.36	3.18

* N = 333

NOR	Min.	Median	Mean	Max.
None	.61	1.31	1.38	2.93
Non-formal	.45	.64	.66	1.25
Formal	.63	1.73	2.00	5.71

* N = 508

POL	Min.	Median	Mean	Max.
None	.04	.65	.61	1.74
Non-formal	.06	.92	.91	2.48
Formal	.22	1.08	1.97	8.23

* N = 361

SVK	Min.	Median	Mean	Max.
None	.22	.72	.78	2.12
Non-formal	.21	.67	.73	1.85
Formal	.53	3.51	3.85	14.69

* N = 485

SWE	Min.	Median	Mean	Max.
None	.47	1.07	1.24	3.03
Non-formal	.27	.66	.70	1.62
Formal	.88	2.36	3.32	11.14

* N = 476

UK	Min.	Median	Mean	Max.
None	.03	.69	.81	4.42
Non-formal	.01	.69	.75	3.33
Formal	.04	1.47	1.77	6.81

* N = 459

USA	Min.	Median	Mean	Max.
None	.44	1.49	1.67	5.79
Non-formal	.28	.77	.82	2.31
Formal	.23	.59	.68	1.99

* N = 430

The next step involves evaluating pairwise covariate balance across the three formats of AET participation. Following the guidelines that considered the standardized effect size of less than .20 as small, .40 as moderate and .60 as large (Cohen, 1988), the results are quite consistent with that for full-time professionals. That is, the maximum standardized effect size between None and Non-formal AET participation is under the moderate threshold (i.e, < .40) in most countries, except for the United Kingdom where the maximum standardized effects size is marginally above .40.

On the other hand, the maximum standardized effect size between None and Formal AET participation exceeds .40 in 9 countries (compared to 7 countries for the sample of full-time professionals) – the Czech Republic (.1.03), Germany (.68), Poland (.63), the United Kingdom (.62), Sweden (.59), South Korea (.57), Belgium (.55), Denmark (.55), Japan (.55). Moreover, the maximum standardized effect size between Non-formal and Formal AET participation exceeds .40 in these same countries with the exception of Poland (.38) and the United Kingdom (.25).

Table 49. Summary of the Standardized Effect Size between Pairs of AET Participation

BEL	Min.	Median	Mean	Max.
Std.eff.sz	.01	.05	.06	.23
Std.eff.sz.1	.01	.13	.18	.55
Std.eff.sz.2	.00	.11	.14	.51

CZE	Min.	Median	Mean	Max.
Std.eff.sz	.00	.05	.07	.33
Std.eff.sz.1	.04	.16	.26	1.04
Std.eff.sz.2	.01	.16	.26	1.18

DNK	Min.	Median	Mean	Max.
Std.eff.sz	.00	.05	.05	.13
Std.eff.sz.1	.00	.16	.16	.55
Std.eff.sz.2	.00	.14	.16	.53

GEM	Min.	Median	Mean	Max.
Std.eff.sz	.00	.03	.04	.13
Std.eff.sz.1	.01	.18	.21	.68
Std.eff.sz.2	.00	.19	.21	.60

IRL	Min.	Median	Mean	Max.
Std.eff.sz	.00	.05	.08	.25
Std.eff.sz.1	.00	.09	.14	.45
Std.eff.sz.2	.00	.07	.09	.37

JPN	Min.	Median	Mean	Max.
Std.eff.sz	.00	.03	.04	.12
Std.eff.sz.1	.00	.10	.17	.55
Std.eff.sz.2	.01	.12	.17	.65

KOR	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.04	.06	.21
Std.eff.sz.1	.00	.23	.24	.57
Std.eff.sz.2	.01	.19	.21	.50

NLD	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.07	.08	.24
Std.eff.sz.1	.02	.17	.18	.35
Std.eff.sz.2	.00	.13	.13	.39

NOR	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.05	.05	.14
Std.eff.sz.1	.00	.08	.09	.27
Std.eff.sz.2	.00	.04	.06	.18

POL	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.06	.07	.26
Std.eff.sz.1	.01	.09	.17	.63
Std.eff.sz.2	.01	.11	.13	.38

SVK	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.03	.04	.11
Std.eff.sz.1	.00	.08	.14	.43
Std.eff.sz.2	.01	.10	.14	.45

SWE	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.01	.05	.05	.12
Std.eff.sz.1	.00	.10	.15	.59
Std.eff.sz.2	.00	.11	.16	.63

UK	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.02	.07	.10	.46
Std.eff.sz.1	.01	.12	.14	.62
Std.eff.sz.2	.01	.08	.08	.25

<u>USA</u>	**Min.**	**Median**	**Mean**	**Max.**
Std.eff.sz	.00	.04	.05	.14
Std.eff.sz.1	.00	.11	.10	.22
Std.eff.sz.2	.01	.10	.11	.29

In short, there is a lack of sufficient covariate balance between None and Formal AET participation and between Non-formal and Formal AET participation for some baseline covariates in the sample of full-time associates. In Tables 50 and 51, we identify covariates did not meet the .40 balance criteria in each of the aforementioned countries for None vs. Formal AET participation and Non-formal vs. Formal AET participation respectively.

Table 50. Countries with Maximum Standardized Effect Sizes > .4 between None and Formal AET Participation (Table 43b, corresponding to Table 43)

Country	Covariates
The Czech Republic	**EDU: B.A.**
Germany	**NUMWORK**
Poland	EDU: B.A.
The United Kingdom	ISCO2C: Health
Sweden	**PARED: at least one secondary**
South Korea	READWORK
Belgium	**EDU: B.A.**
Denmark	**INFLUENCE**
Japan	**WRITEWORK**

** Covariates that did not meet the .40 balance criteria for both None vs. Formal AET participation and Non-formal vs. Formal AET participation are bolded.*

Table 51. Countries with Maximum Standardized Effect Sizes > .4 between Non-formal and Formal AET Participation (Table 44b, corresponding to Table 44)

Country	Covariates
The Czech Republic	**EDU: B.A.** ISCO2C: Health
Germany	**NUMWORK**
Sweden	**PARED: at least one secondary**
South Korea	INFLUENCE
Belgium	**EDU: B.A.**
Denmark	**INFLUENCE**
Japan	**WRITEWORK**

** Covariates that did not meet the .40 balance criteria for both None vs. Formal AET participation and Non-formal vs. Formal AET participation are bolded.*

A comparison of Tables 43 and 44 (for full-time professionals) with Tables 50 and 51 (for full-time associates) reveals that the Czech Republic, Germany and Japan are the three countries with maximum standardized effect sizes greater than .40 in both samples. For the sample of full-time associates, the Czech Republic is the country with the largest maximum effect sizes for both None vs. Formal AET participation (1.04) and Non-formal vs. Formal AET participation (1.18). The covariate EDU: B.A. failed to reach the .40 balance criteria across samples. For the sample of full-time professionals, Japan is the country with the largest maximum effect sizes for both None vs. Formal AET participation (.80) and Non-formal vs. Formal AET participation (.86). And it is the covariate WRITWORK that did reach the .40 balance criteria across samples.

As suggested at the end of Section 4.3.4, covariates that did not achieve balance within .40 standard deviations after weighting will be included in the outcome model to increase the robustness of the ATE estimators to misspecifications of the propensity score model (i.e., doubly robust estimation). Weighted regression on treatment indicators and unbalanced covariates (including interactions between unbalanced covariates and the treatment condition) for full-time associates yields results comparable to that for full-time professionals (please refer back to Tables 50 and 51 for more information).

Table 52. Estimated ATEs of the Different Formats of AET Participation on Full-time Associates' PS-TRE Scores (Table 45b, corresponding to Table 45)

BEL	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	1.14	6.80	.17	.87
Formal vs. None	9.71	13.74	.71	.48
Non-formal vs. Formal	-8.57	13.34	-.64	.52

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

CZE	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	-9.21	17.31	-.53	.60
Formal vs. None	38.80	45.87	.85	.40
Non-formal vs. Formal	-48.01	44.75	-1.07	.28

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

DNK	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	12.20	16.04	.76	.45
Formal vs. None	31.30	23.56	1.33	.18
Non-formal vs. Formal	-19.10	22.82	-.84	.40

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

GEM	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	8.26	12.39	.67	.51
Formal vs. None	-2.56	25.32	-.11	.92
Non-formal vs. Formal	10.82	24.17	.45	.65

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

IRL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	11.71	23.34	-.50	.62
Formal vs. None	10.60	30.79	.34	.73
Non-formal vs. Formal	-22.31	24.62	-.91	.37

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	9.11	12.92	.70	.48
Formal vs. None	.87	35.56	.02	.98
Non-formal vs. Formal	8.23	34.24	.24	.81

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

KOR	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-12.53	17.66	-.71	.48
Formal vs. None	-26.39	47.73	-.55	.58
Non-formal vs. Formal	13.86	49.32	.28	.78

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

NLD	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-25.51	22.06	-1.16	.25
Formal vs. None	14.07	19.47	.72	.47
Non-formal vs. Formal	-38.58	20.73	-1.91	.06

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

NOR	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	25.22	18.62	1.35	.18
Formal vs. None	66.48	27.54	2.41	.02
Non-formal vs. Formal	-41.26	25.48	-1.62	.11

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

POL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	31.95	22.91	1.39	.16
Formal vs. None	-19.51	26.73	-.73	.47
Non-formal vs. Formal	51.46	29.32	1.76	.08

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SVK	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-2.95	28.84	-.10	.92
Formal vs. None	-19.01	35.80	-.53	.60
Non-formal vs. Formal	16.06	25.10	.64	.52

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SWE	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	1.14	6.80	.17	.87
Formal vs. None	9.71	13.74	.71	.48
Non-formal vs. Formal	-8.57	13.34	-.64	.52

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

UK	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	9.83	28.43	.35	.73
Formal vs. None	-36.80	30.83	-1.19	.23
Non-formal vs. Formal	46.63	24.13	1.93	.05

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

USA	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-14.90	23.49	-.63	.53
Formal vs. None	68.67	25.20	2.73	.01**
Non-formal vs. Formal	-83.56	26.29	-3.18	.0012**

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

The first thing that stands out is that the United States is the only country where Formal AET participation is significantly and positively associated with full-time associates' PS-TRE proficiency scores. Compared to non-participants with similar propensity scores for Formal AET participation, Formal AET participants are estimated to score 68.67 points higher on PS-TRE ($p = .01$). Compared to Non-formal AET participants with similar propensity scores for Non-formal

AET participation, Formal AET participants are estimated to score 83.56 points higher on PS-TRE ($p < .01$). As for the binary outcome, the probabilities of scoring in the top quartile of the PS-TRE score distribution for Formal AET participants are significantly higher than Non-formal AET participants with similar propensity scores for Non-formal AET participation ($p < .01$).

Recall that in the sample of full-time professionals, Denmark is the only country where both formats of AET participation (vs. None) are significantly and positively associated with PS-TRE proficiency scores and the probability of scoring in the top quartile of the PS-TRE score distribution. However, AET participation (either Formal or Non-formal) is only significantly associated with full-time associates' probability of scoring in the top quartile in Denmark. Specifically, the probabilities of scoring in the top quartile of the PS-TRE score distribution for Formal or Non-formal AET participants are significantly higher than non-participants with similar propensity scores for Formal or Non-formal AET participation ($p < .01$). Among those with similar propensity scores for Non-formal AET participation, the probabilities of scoring in the top quartile of the PS-TRE score distribution for Formal AET participants are significantly higher than that for Non-formal AET participants ($p < .01$).

Lastly, there are three more countries where Formal AET participation seems to have a negative association with full-time associates' probability of scoring in the top quartile – Japan, Poland and Sweden. However, as mentioned in Section 4.3.5, these results ought to be treated with extreme caution due to the relatively few Formal AET participants (i.e., unbalanced class) in these countries. Please refer to Figure 37 (or Figure 33b) and Table 53 (or Table 46b) for more information.

Table 53. Estimated ATEs of the Different Formats of AET Participation on Full-time Associates' Probability of Scoring in the Top Quartile of the PS-TRE Score Distribution (Table 46b, corresponding to Table 46)

DNK	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	9.51	3.15	3.02	.002**
Formal vs. None	29.82	5.37	5.56	.000***
Non-formal vs. Formal	-20.32	4.59	-4.43	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	1.11	1.09	1.01	.31
Formal vs. None	-13.18	2.55	-5.17	.000***
Non-formal vs. Formal	14.29	2.44	5.86	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

POL	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	1.83	2.70	.68	.51
Formal vs. None	-27.04	5.92	-4.57	.000***
Non-formal vs. Formal	28.87	5.80	4.98	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SWE	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	5.26	5.86	.90	.37
Formal vs. None	-111.75	6.98	-16.01	.000***
Non-formal vs. Formal	117.01	4.04	28.99	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

USA	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	-2.49	2.44	-1.02	.31
Formal vs. None	5.86	2.52	2.33	.02*
Non-formal vs. Formal	-8.35	2.41	-3.46	.001***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

4.4.2 Parallel Analyses with Full-time Professionals' Literacy and Numeracy Proficiency Scores as Outcome Variables

Instead of PS-TRE, the second set of parallel analyses focus on full-time professionals' proficiency in traditional cognitive skills (i.e., Literacy and Numeracy). In accordance with the OECD description that "high levels of proficiency in literacy and numeracy go hand in hand with high levels of proficiency in problem solving in digital environments" (OECD, 2013c, p. 96), RQ1c finds that among full-time professionals from the 14 selected countries, the correlation between proficiency scores in PS-TRE and Literacy is .79, while the correlation between PS-TRE and Numeracy is .74. What is more, the correlation between proficiency scores in Literacy and Numeracy is .84 for the sample of full-time professionals, compared with .87 for the entire PIAAC sample. Please refer to Section 4.1.3 for detailed descriptions of the correlations across countries.

However, different results emerge when we estimate the associations between the three formats of AET participation and full-time professionals' Literacy and Numeracy proficiency scores respectively. Recall that in the main analyses (i.e., RQ3), both formats of AET participation (vs. None) are significantly and positively associated with full-time professionals' PS-TRE proficiency scores and their probability of scoring in the top quartile of the PS-TRE score distribution in Denmark. When it comes to full-time professionals' Literacy proficiency scores, the associations are not significant at the .01 level in all 14 countries.

With regard to full-time professionals' probability of scoring in the top quartile of the Literacy distribution, Poland and Sweden are the two countries where AET participation (either Formal or Non-formal) is significantly associated with an increased probability of scoring at or above the 75th percentile of the Literacy distribution ($p < .01$), controlling for propensity scores for Formal or Non-formal AET participation, unbalanced covariates and interactions between unbalanced covariates and the treatment condition. It is also interesting to note that the Slovak Republic is the only country where Formal AET participation (vs. None) is significantly and negatively associated with full-time professionals' probability of scoring in the top quartile of the Literacy distribution ($p < .01$). That is, the probabilities of scoring in the top quartile of the Literacy distribution for Formal AET participants are significantly lower than non-participants with similar scores for Formal AET participation ($p < .01$). Among those with similar propensity scores for Non-formal AET participation, the probabilities of scoring in the top quartile of the PS-TRE score distribution for Formal AET participants are significantly lower than that for Non-formal AET participants ($p < .01$). Please see Table 54 below for more information.

Table 54. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the LIT Distribution

POL	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	17.14	3.29	5.21	.000***
Formal vs. None	14.20	4.87	2.92	.004**
Non-formal vs. Formal	2.94	4.06	.72	.47

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SVK	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	.31	.83	.37	.71
Formal vs. None	-17.15	1.18	-15.52	.000***
Non-formal vs. Formal	17.45	1.10	15.93	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

SWE	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	72.35	24.62	2.94	.003**
Formal vs. None	76.59	24.66	3.11	.002**
Non-formal vs. Formal	-4.23	2.11	-2.01	.05

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

When it comes to full-time professionals' Numeracy proficiency scores, Ireland is the only country where both formats of AET participation (vs. None) are significantly and positively associated with Numeracy proficiency scores ($p < .01$). What is more, the probabilities of scoring in the top quartile of the Numeracy distribution for Non-formal AET participants are significantly higher than non-participants with similar propensity scores for Non-formal AET participation ($p < .01$). Please see Tables 56 and 57 below for detailed information. Ireland is comparable to Denmark where both formats of AET participation (vs. None) are significantly and positively associated with full-time professionals' PS-TRE proficiency scores and their probability of scoring in the top quartile of the PS-TRE score distribution.

Next, Japan and Poland are the two countries where Formal AET participants are estimated to score significantly higher than Non-formal AET participants with similar propensity scores for Non-formal AET participation ($p < .01$). What is more, the probabilities of scoring in the top quartile of the Numeracy distribution for Formal AET participants are significantly higher than non-participants with similar propensity scores for Formal AET participation ($p < .01$). However, as mentioned in Section 4.3.5, the results from Japan and Poland ought to be treated with extreme caution due to the relatively few Formal AET participants (i.e., unbalanced class). Please refer to Figure 33 and Tables 55 and 56 below for more information.

Table 55. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' NUM Proficiency Scores

IRL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	92.44	28.45	3.25	.001***
Formal vs. None	85.87	31.73	2.71	.01**
Non-formal vs. Formal	6.56	24.34	.27	.79

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	16.46	15.86	-1.04	.30
Formal vs. None	66.49	31.95	2.08	.04
Non-formal vs. Formal	-82.96	30.74	-2.70	.01**

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

POL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	-42.01	34.59	-1.21	.23
Formal vs. None	99.41	38.94	2.55	.01**
Non-formal vs. Formal	-141.42	39.31	-3.60	.0004***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

Table 56. Estimated ATEs of the Different Formats of AET Participation on Full-time Professionals' Probability of Scoring in the Top Quartile of the NUM Distribution

IRL	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	11.22	3.12	3.60	.0004***
Formal vs. None	7.96	3.57	2.23	.03
Non-formal vs. Formal	3.26	2.65	1.23	.22

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

JPN	Estimate	Std. Error	t value	Pr(>|t|)
Non-formal vs. None	2.31	1.89	1.22	.22
Formal vs. None	28.05	5.92	4.74	.000***
Non-formal vs. Formal	-25.74	5.77	-4.46	.000***

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

<u>**POL**</u>	**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**
Non-formal vs. None	3.95	571.69	.01	.99
Formal vs. None	26.69	3.62	7.38	.000***
Non-formal vs. Formal	-22.75	571.69	-.04	.97

*** significant at .001 level, ** significant at .01 level, * significant at .05 level (2-tailed)

Chapter 5. Conclusions

5.1 Overview of Research Questions

In today's digital world, good problem solving skills are critical to both organizations' and individuals' success in an increasingly competitive global market. Unlike traditional cognitive skills (i.e., Literacy and Numeracy) which are typically taught as school subjects, problem solving skills, for the most part, can be developed through work-related experience and adult education and training initiatives (Sgobbi, 2014). What is more, because "performance in problem solving is less dependent on achievement and opportunities of early age", AET programs may serve as a remedy for educational inequity because of a disadvantaged family background (Cegolon, 2016).

This study addressed the key question regarding the relationships between different formats of AET participation and full-time professionals' PS-TRE proficiency (as measured by proficiency scores and their probabilities of scoring in the top quartile of the PS-TRE distribution). To test the robustness of the estimates with different populations and different skill domains, two sets of parallel analyses were conducted: one estimates the associations between different formats of AET participation and full-time associates' PS-TRE proficiency and the other estimates the associations between different formats of AET participation and full-time professionals' Literacy and Numeracy proficiencies.

As preparation for the key question (i.e., RQ3), RQ1 provided detailed examination of the outcome variable: full-time professionals' PS-TRE proficiency scores. First of all, the distributions of full-time professionals' PS-TRE proficiency scores were compared across the 14 selected countries. Within each country, the distributions were further grouped by gender and age. Patterns observed from the overall and gender-/age-specific distributions were compared across countries. Please see Section 5.2 for a brief summary of findings.

Because the first set of parallel analyses were introduced under the hypothesis that PS-TRE proficiency should “go hand in hand” with Literacy and Numeracy proficiency (OECD, 2013c), the correlation between full-time professionals’ PS-TRE proficiency scores and their Literacy and Numeracy proficiency scores were calculated and compared across countries. On the other hand, the second set of parallel analyses were introduced based on the understanding that full-time professionals should possess higher PS-TRE proficiency than other less privileged groups in terms of educational attainment and social capital. Therefore, RQ1 also looked into the distributions of full-time associates’ PS-TRE proficiency scores and compared the results to that of full-time professionals.

RQ2 set the path to RQ3 by giving us a picture of significant predictors of full-time professionals’ probabilities of AET participation in each of the 14 selected countries. Since full-time professionals can participate in Formal, Non-formal or None AET programs, predictors for different formats of AET participation were reported separately. Comparing different patterns for Formal/Non-formal AET participation (vs. None) across the 14 selected countries helped us understand some systematic differences that already existed before AET participation.

The purpose of RQ3 is to estimate the associations between different formats of AET participation and full-time professionals’ PS-TRE proficiency. Propensity score methods were deployed to make sure that we are comparing Formal/Non-formal AET participants to Non-participants with similar probabilities of a certain format of AET participation. Assuming all confounding variables were properly controlled, the differences we observed in the outcomes (i.e., PS-TRE proficiency scores or the probabilities of scoring in the top quartile of the PS-TRE distribution) can be attributed to the specific format of AET participation.

However, given the nature of observational studies, we can only control for observed covariates (and the components of unobserved covariates that are correlated with the observed covariates). Consequently, any conclusions drawn from this study must be considered provisional. As will be discussed in Section 5.2, the significant associations observed between AET participation (vs. None) and (1) full-time professionals' PS-TRE proficiency in Denmark, (2) full-time associates' PS-TRE proficiency in the United States, and (3) full-time professionals' Numeracy proficiency in Ireland are subject to the hidden bias due to unobserved confounders that are not correlated with the observed covariates but are correlated with the outcome.

To test the robustness of the estimated associations between full-time professionals' PS-TRE proficiency and different formats of AET participation, two sets of parallel analyses were conducted in RQ4. As in the main analyses, the lack of systematic patterns of association was consistent across different populations and different skill domains, with the exception of a few countries. Section 5.2 highlights these countries and then discusses potential directions for future studies.

5.2 Summary of Findings by Research Question

Research Questions 1a & 1b

On average, full-time professionals (at least 25 years old) in all the 14 selected countries performed at Level 2 (scores from 291 to 340) in PS-TRE. Sweden, the Netherlands and Japan are countries with the highest mean PS-TRE scores while Poland, Ireland, the Slovak Republic and South Korea are countries with the lowest mean PS-TRE scores. The rankings of countries did not change much when comparing the 80th, 85th and 90th percentiles of the score distribution for each country.

When splitting the sample of full-time professionals by gender, males scored higher than females in all 14 countries. However, Japan's ranking dropped for females but remained on top for males. In fact, the gender gap in Japan is almost as large as that in Poland – the country with the lowest mean PS-TRE scores and the largest gender gap favoring males. On the other hand, the smallest gender gaps in favor of males are observed in South Korea and the Czech Republic. These findings suggested that the two East Asian countries are different not only in terms of PS-TRE proficiency scores but in the extent of gender gap. In a similar fashion, the Czech Republic may be different from the other two Visegrád Group countries with low mean PS-TRE scores but large gender gaps.

Splitting the sample by age cohort manifested an age-related decline in PS-TRE proficiency scores as measure by PIAAC. The mean scores tended to decrease from younger to older cohorts in all 14 countries. Again, Japan is the country with the largest age-related gap in PS-TRE proficiency scores -- while the oldest cohort (i.e., 55-65 year olds) scored relatively low on PS-TRE, the youngest cohort (i.e., 25-34 year olds) scored only second to Sweden. The United States is quite different from Japan in that the younger generation (i.e., 25-34 year olds) scored more or less the same as the older generation (i.e., 55-65 year olds), and, thus, displays the smallest gap while other countries demonstrate generally steady progress across generations. Further splitting the sample by gender and age cohort simultaneously confirmed that the greatest gender*age gap was between the youngest male and the oldest female subgroups in Japan, while the smallest was observed in the United States.

Research Question 1c

Because the PS-TRE assessment of PIAAC focuses on proficiency in Literacy and technology (OECD, 2013c), the correlation between PS-TRE and Literacy proficiency scores at the individual level is somewhat higher than that between PS-TRE and Numeracy proficiency scores (i.e., .79 > .74). If the magnitudes of such correlations indicates to degree to which "the digital divide reflecting a Literacy divide", the results suggest that full-time professionals' PS-TRE proficiency is highly correlated with Literacy proficiency in the three Scandinavian countries, the Netherlands, the United Kingdom and the United States, with lower correlations in the four Central European and the two East Asian countries.

Research Question 1d

Full-time associates (at least 25 years old) scored lower than full-time professionals in all 14 countries, while 10 of them maintained their average performance at Level 2. Japan, Sweden and Norway are countries with the highest mean PS-TRE proficiency scores while Poland, Ireland, the Slovak Republic and South Korea are countries with the lowest mean PS-TRE scores. The rankings of countries are almost identical to those for full-time professionals except that the United States dropped from the 10^{th} to the very bottom of the list. The ranking of Another difference is the number of countries performing at Level 3 at the 80^{th}, 85^{th} and 90^{th} percentiles of the score distribution – one, one and five – compared with two, six and eleven in the sample of full-time professionals.

When splitting the associate professionals sample by gender, males scored higher than females in 12 countries. Japan's ranking remained on top for both genders. However, Japanese females scored 6 points higher than males, which is the greatest gender gap favoring females in the sample of full-time associates. Recall that Japan is the country with the greatest gender gap favoring males in the sample of full-time professionals. Such a dramatic contrast hints at the existence of systematic differences between the two populations. To identify possible explanations for such a shift, a more intense investigation of employment patterns of the Japanese labor market is needed.

Splitting the associated sample by age cohort yielded results similar to what was found with the sample of full-time professionals. The age-related gap in PS-RE proficiency scores is the largest in Japan and the smallest in the United States. Such findings are particularly concerning for the 25-34-year-old Americans given that their peers in all the other 13 countries have shown larger gains (i.e., positive differences) in PS-TRE proficiency scores over the oldest cohort. Splitting the sample further by gender and age cohort simultaneously only reinforced the concern because for either gender, the United States is the only country where the youngest cohort did not score higher than the older ones.

Lastly, the correlation between PS-TRE and Literacy proficiency scores at the individual level is also higher than that between PS-TRE and Numeracy proficiency scores for full-time associates (i.e., .80 > .75). The cross-country distribution of correlation coefficients follows a similar pattern to that of full-time professionals except for Sweden. Despite a high correlation between full-time professionals' PS-TRE and Literacy proficiencies in Sweden, the correlation is much smaller when it comes to full-time associates. Again, such change suggests that the two samples might be different in a systematic way.

Research Question 2

Research question two modelled full-time professionals' probabilities of participating in the different formats of AET programs across the 14 selected countries. Since parameter estimates were reported for comparisons of Non-formal (vs. None) and Formal (vs. None) AET participation respectively, unique patterns can be summarized by examining the similarities and differences between the two sets of results.

In comparing significant predictors for Non-formal and Formal AET participation, the first thing that stands out is that across countries, age cohorts play an important role in predicting full-time professionals' probability of Formal AET participation in up to four countries, but not so much in their probability of Non-formal AET participation. Specifically, in Belgium, Denmark, Sweden and the United States, the youngest age cohort (i.e., 25-34 year olds) is associated with a higher probability of Formal AET participation than the oldest age cohort ((i.e., 55-65 year olds), after partialling out the variability due to the other variables in the model.

Secondly, in the two East Asian countries, a higher level of educational attainment is significantly related to a higher probability of participating in both AET programs vs. None. After partialling out the variability due to the other variables in the model, the relationships between educational attainment and Formal AET participation is stronger than that between educational attainment and Non-formal AET participation. Please refer to Tables 18 and 19 in Section 4.2.1 and Tables 30 and 31 in Section 4.2.2 for more information.

Thirdly, after partialling out the variability due to the other variables in the model, Health professionals are estimated to have a significantly higher probability of participating in both formats AET programs in Belgium and the Slovak Republic ($p < .01$). Compared to their peers in Legal, Social and Cultural fields, Teaching professionals are estimated to have a higher probability for Non-formal AET participation in Ireland but a higher probability for Formal AET participation in the Slovak Republic.

The fourth finding concerns the relationships between full-time professionals' probability of AET participation vs. None and their use of different domains of skills. As is evident from

Sections 4.2.1 and 4.2.2, the Netherlands and the United Kingdom are the two countries where indicators of skills use significantly predict both formats of AET participation (vs. None). For Dutch full-time professionals, after partialling out the variability due to the other variables in the model, a one-unit increase in the use of Writing skills at work is associated with a 130% increase in the odds of Formal AET participation since exp(.83) = 2.30, while a one-unit increase in the use of Numeracy skills at work is associated with a 39% decrease in the odds of Non-formal AET participation since exp(-.50) = .61 (see Tables 22 and 34). For British full-time professionals, after partialling out the variability due to the other variables in the model, a one-unit increase in the use of Influence skills at work is associated with a 50% increase in the odds of Non-formal AET participation and a 58% increase in the odds of Formal AET participation. Also note that everything else being equal, a one-unit increase in the use of Task Discretion skills in the workplace is associated with a 36% decrease in British full-time professionals' odds of Formal AET participation (see Tables 23 and 36).

Research Questions 3 & 4

The results from research question three and its parallel analyses (i.e., research question four) were already summarized in Chapter 4 based on the different outcomes and samples. After controlling for other variables through propensity score weighting, there are only scattered significant findings across the 14 selected countries. Here, we re-organize the results by the combination of outcome and sample and compare the different patterns observed across countries. Table 58 summarizes estimated ATEs of different formats of AET participation on different outcomes and different samples by country. Note that countries where AET participation (either Formal or Non-formal) is significantly associated with both proficiency scores and the probability of scoring in the top quartile are bolded.

Table 57. Summary of Estimated ATEs of the Different Formats of AET Participation on the Different Outcomes and Different Samples by Country

Full-time Professionals' Problem-Solving Scores			
Non-formal vs. None	**DNK**		
Formal vs. None	**DNK**		
Non-formal vs. Formal			
Full-time Professionals' Problem-Solving Top Quartile			
Non-formal vs. None	**DNK**		
Formal vs. None	**DNK**		
Non-formal vs. Formal			
Full-time Associates' Problem-Solving Scores			
Non-formal vs. None			
Formal vs. None		**USA**	
Non-formal vs. Formal		**USA**	
Full-time Associates' Problem-Solving Top Quartile			
Non-formal vs. None	DNK		
Formal vs. None	DNK		
Non-formal vs. Formal	DNK	**USA**	
Full-time Professionals' Numeracy Scores			
Non-formal vs. None			**IRL**
Formal vs. None			**IRL**
Non-formal vs. Formal			
Full-time Professionals' Numeracy Top Quartile			
Non-formal vs. None			**IRL**
Formal vs. None			
Non-formal vs. Formal			

Starting from the main analyses (i.e., RQ3), perhaps the most prominent finding is the close relationships between both formats of AET participation (vs. None) and full-time professionals' PS-TRE proficiency in Denmark. Specifically, participating in AET programs (either Formal or Non-formal) is significantly and positively associated with Danish full-time professionals' PS-TRE proficiency scores and their probability of scoring in the top quartile of the PS-TRE score distribution ($p < .01$ for one significance test conducted). For Danish full-time associates, participating in AET programs (either Formal or Non-formal) is also significantly and positively associated with their probability of scoring in the top quartile of the PS-TRE score distribution ($p < .01$ for one significance test conducted).

By contrast, there is no significant relationship at the .01 level between AET participation (either Formal or Non-formal) and full-time professionals' PS-TRE proficiency in the United States[24]. The same lack of significant results at the .01 level was also observed between AET participation (either Formal or Non-formal) and full-time professionals' Literacy and Numeracy proficiency in the United States. However, when it comes to full-time associates, Formal AET participants are estimated to score 69 points higher on PS-TRE ($p = .01$) than non-participants with similar propensity scores for Formal AET participation and 84 points higher on PS-TRE ($p < .01$) than Non-formal AET participants with similar propensity scores for Non-formal AET participation ($p < .01$). Moreover, the probabilities of scoring in the top quartile of the PS-TRE score distribution for Formal AET participants are significantly higher than Non-formal AET participants with similar propensity scores for Non-formal AET participation ($p < .01$). In short, Formal AET programs in the United States play a significant role in developing and improving PS-TRE proficiency for full-time associates, but not for full-time professionals.

[24] Three significance tests were conducted due to the existence of one unbalanced covariate (i.e., EDU: less than B.A.) in the outcome model for the United States.

As for the associations between full-time professionals' AET participation and their Literacy proficiency scores, no significant results are observed at the .01 level in all 14 countries. The results support Sgobbi's (2014) view that Literacy, for the most part, are fostered through formal education and is less dependent on work-related experience and adult education. When it comes to full-time professionals' Numeracy proficiency scores, the most important finding is that in Ireland, AET participants (either Formal or Non-formal) are estimated to score significantly higher on Numeracy than non-participants with similar propensity scores for Formal or Non-formal AET participation ($p < .01$ for one significance test conducted). What is more, the probabilities of scoring in the top quartile of the Numeracy distribution for Non-formal AET participants are significantly higher than non-participants with similar propensity scores for Non-formal AET participation ($p < .01$ for one significance test conducted). Such relationships are comparable to the associations between AET participation (either Formal or Non-formal) and full-time professionals' PS-TRE proficiency in Denmark, and the associations between AET participation (either Formal or Non-formal) and full-time associates' PS-TRE proficiency in the United States.

5.3 Limitations and Future Work

The variations in relationships between the different formats of AET participation and working adults' skills proficiency across domains and samples indicate the necessity of conducting qualitative research on AET programs in individual countries. This is particularly true given the cross-sectional nature of the PIAAC data. As stated in Section 1.6, it is very difficult to draw credible causal inferences from observational studies. Of particular concern is the potential existence of unobserved confounders that may undermine the validity of causal inferences from observational studies. For example, it is acknowledged that different national institutional settings should play an important role in Formal and Non-formal AET participation rates. However, it is hard to take into account all the features of a national institutional setting (i.e., unobserved confounders). According to Cegolon (2016), countries with different national institutional settings can differ considerably with respect to the characteristics of their educational, training, and occupational system, their labor market regulations, the nature of their employment-sustaining policies and the national welfare systems. Therefore, it is desirable to conduct qualitative research on AET programs in individual countries and then provide recommendations tailored to the specific needs of each country.

Like other conventional regression methods, propensity score methods can account for only observed confounders (and the components of unobserved confounders correlated with the observed ones) but not unobserved confounders that are not correlated with the observed ones. Although the combination of propensity score methods and parallel analyses helps boost our confidence in the rigor and robustness of the obtained estimates, they cannot fully eliminate hidden bias due to due to unobserved confounders that are not correlated with the observed covariates but are correlated with the outcome.

To correct for unobserved confounders in observational studies, the instrument variables (IV) method has been proposed to identify the unobserved correlation between the explanatory variable and response variable. However, its use in observational studies is limited because appropriate instruments are hard to found. What is more, it was found that instrumental variables are less useful in case of strong confounding because a strong instrument cannot be found and the usual assumptions will be easily violated (Martens et al., 2006).

Alternatively, sensitivity analyses can be used to assess the drawn conclusions' degree of robustness to hidden bias due to unobserved confounders that are not correlated with the observed covariates but are correlated with the outcome. More specifically, sensitivity analyses examine “how strong the correlations would have to be between a hypothetical unobserved covariate and both treatment assignment and the outcome to make the observed treatment effect go away” (Rosenbaum & Rubin, 1983; Rosenbaum, 2005). Although sensitivity analysis methods are becoming more and more developed, they are still used relatively infrequently (Stuart, 2010). Hopefully, more software will be available to help their adoption by more researchers.

What is more, high-dimensional propensity score has been suggested to reduce the risk of unobserved confounders. It is based on the understanding that propensity score analyses are only as good as the completeness and quality of the potential confounders available (Ali et al., 2019). Constructing the propensity score model with a rich set of covariates (up to several hundred, if possible) could potentially increase the probability that at least some of the observed covariates may collectively be proxies for unobserved confounders (Schneeweiss et al., 2009; Rassen et al., 2011). Therefore, it is necessary to conduct a mixed-method approach (e.g., combining propensity score models with policy analyses for individual countries) that not only puts the high-dimensional data into proper context but also prevents the misinterpretation of statistical models.

Last but not least, even within the same country, AET programs can differ substantially in terms of their objectives, structures, content and levels. This is particularly true when it comes to Non-formal AET programs. Take Denmark as an example, popular adult education programs offer general competences up to upper-secondary level while labor market education programs provide vocational training up to skilled level (Danish Adult Education Association, 2019).

Future studies of adult education may benefit from a fine-grained classification of AET programs based on the guideline provided by OECD (2001):

- institutionalized, school-based education
- institutionalized, job-related training
- planned and structured in-house learning activities (on the job learning)
- non-formal in-house learning activities (in the job learning)

References

Agresti, A. (2002). *Categorical Data Analysis* (2^{nd} ed.). Hoboken, NJ: John Wiley.

Ali, M., Prieto-Alhambra, D., Lopes, L., Ramos, D., Bispo, N., Ichihara, M., Pescarini, J., Williamson, E., Fiaccone R., Barreto, M., & Smeeth, L. (2019). Propensity Score Methods in Health Technology Assessment: Principles, Extended Applications, and Recent Advances. *Front. Pharmacol., 10*(973).

Allison, P. (2004). Convergence problems in logistic regression. In Altman, M., Gill, J., & McDonald, M. (Eds.), *Numerical Issues in Statistical Computing for the Social Scientist.* New York, NY: John Wiley, 238-252.

Allison, P. (2014). Measures of fit for logistic regression. Paper 1485-2014, SAS Global Forum, 2014, Washington, D.C..

American Management Association. (2010). *AMA 2010 critical skills survey. Executive summary.* Retrieved from http://www.amanet.org/training/articles/3727.aspx

Anglemyer, A., Horvath, H., & Bero, L. (2014) Healthcare outcomes assess with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev, 29*(4).

Arora, A., & Pawlowksi, E. (2017). *Examining gender differences in the mathematical literacy of 15-year-olds and the numeracy skills of the age cohorts as adults.* Retrieved from https://static1.squarespace.com/static/51bb74b8e4b0139570ddf020/t/588a18c13a04118a0c68116e/1485445313820/Arora_Pawlowski_PIAAC_2017.pdf

Austin, P. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46,* 399-424.

Austin, P. (2019). Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Stat Methods Med Res, 28*(5), 1365-1377.

Autor, D., Levy F., & Murnane R. (2003). The skill content of recent technical change: An empirical investigation. *Quarterly Journal of Economics, 118*(4), 1279-1334.

Bang, H. & Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics, 61,* 962–972.

Bassanini, A., Alison, B., Brunello, G., de Paola, M., & Leuven, E. (2007). Workplace training in Europe, In: Brunello, G., Garibaldi, P., & Wasmer, E. (Eds.), *Education and Training in Europe.* Oxford University Press, Chap. 8-13.

Bothwell, L., Greene, J., Podolsky, S., & Jones, D. (2016). Assessing the gold standard – lessons from the history of RCTs. *The New England Journal of Medicine, 374,* 2175-2181.

Bothwell, L., & Podolsky, S. (2016). The emergence of the randomized, controlled trial. *The New England Journal of Medicine, 375,* 501-504.

Braun, H. (2018). How long is the shadow? The relationships of family background to selected adult outcomes: Results from PIAAC. *Large-scale Assess, Educ 6*(4). doi:10.1186/s40536-018-0058-x

Bresnahan, T. (1999). Computerization and wage dispersion: An analytical reinterpretation. *The Economic Journal, 109*(456), 390– 415.

Burgett, L., Griffin, B., & McCaffrey, D. (2017). *Propensity scores for multiple treatments: A tutorial for the mnps function in the twang package.* Retrieved from https://cran.r-project.org/web/packages/twang/vignettes/mnps.pdf

Casner-Lotto, J., & Barrington, L. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st century U.S. workforce*. New York, NY: The Conference Board, Inc.

Cegolon, A. (2016). *Determinants and learning effects of adult education-training: A cross-national comparison using PIAAC data* (Department of Quantitative Social Science Working Paper No. 15-11). UCL Institute of Education, University College London, UK.

Chetty, R., Hendren, N., & Katz, L. (2015). *The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment.* Cambridge, MA: National Bureau of Economic Research. Retrieved from https://www.ets.org/s/research/report/opportunity/ets-choosing-our-future.pdf

Clay, R. (2010). More than one way to measure. *Monitor on Psychology, 41*(8), 52.

Cochran, W. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24,* 295-313.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* New Jersey: Lawrence Erlbaum

Cole, S., & Hernan, M. (2008). Constructing inverse probability weights for marginal structural models. *Am J Epidemiol, 168*(6), 656-65.

Cook, T., & Campbell, D. (1979). *Quasi-Experimentation: Design & Analysis Issues in Field Settings*. Boston, MA: Houghton Mifflin.

Cox. D. & Snell E. (1989). *Analysis of Binary Data* (2nd ed.). Chapman & Hall.

Diamond, A., & Sekhon, J. (2013). Genetic matching for estimating causal effect: A general multivariate matching method for achieving balance in observational studies. *The Review of Economics and Statistics, 95,* 932-945.

Dickerson, A., & Green, F. (2004). The growth and valuation of computing and other generic skills. *Oxford Economic Papers-New Series, 56*(3), 371-406.

Dieckhoff, M., & Steiber, N. (2011). A re-assessement of common theoretical approaches to explain gender differences in continuing training participation. *British Journal of Industrial Relations, 49*(s1), 135-157.

Fisher, R. (1925). *Statistical Methods for Research Workers* (1st ed.). Oliver and Boyd, Edomburgh.

Funk, M., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology, 173*(7), 761–767.

Ginsburg, A., & Smith, M. (2016). *Do Randomized Controlled Trials Meet the 'Gold Standard'?*. American Enterprise Institute.

Goodman, M., Sands, A., & Coley, R. (2015). *America's Skills Challenge: Millennials and the Future*. Princeton: Educational Testing Service.

Gu X., & Rosenbaum P. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics, 2,* 405–420.

Guo, S., & Fraser, M. (2015). *Propensity Score Analysis: Statistical Methods and Applications.* Thousand Oaks, CA: Sage Publications, Inc.

Hanushek E., Woessmann L., & Zhang L. (2011). *General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle* (IZA Discussion Paper No. 6083) Bonn: IZA.

Haneuse, S. (2016). Distinguishing selection bias and confounding bias in comparative effectiveness research. *Med Care, 54*(4), 23-39.

Ho, D., Imai, K., King, G., & Stuart, E. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15,* 199–236.

Hodge, K. & Lear, J. (2011). Employment skills for 21st century workplace: The gap between faculty and student perceptions. *Journal of Career and Technical Education, 26*(2), 28-41.

Holland, P. (1986). Statistics and causal inference. *Journal of the American Statistical Association, 81,* 945–960.

Holland, P. (1988). Causal inference, path analysis and recursive structural equations models. *Sociological Methodology, 18,* 449-484.

Holzer, H. & Lerman, R. (2015). *Cognitive skills in the U.S. labor market: For whom do they matter?* Washington, D.C.: U.S. Program for the assessment of adult competencies. Retrieved from http://static1.squarespace.com/static/51bb74b8e4b0139570ddf020/t/54da7582e4b0cb4c49fc0d9d/1423603074704/Holzer_Lerman_PIAAC.pdf

Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*(3), 706–710.

Imbens, G. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics & Statistics, 86*(1), 4-29.

Imbens, G., & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5-86.

Ioannidis, J., Haidich, A., Pappa, M., Pantazis, N., Kokori, S., Tektonidou, M., Contopoulos-Ioannidis D. & Lau, J. (2001) Comparision of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*, *286*(7):821-30.

Kelley, K., & Preacher, K. (2012). On effect size. *Psychological Methods, 17*(2), 137-152.

Kirsch, I., Braun, H., Lennon, M., & Sands, A. (2016). *Choosing our future: A story of opportunity in America.* Princeton: Educational Testing Service.

Lazonder, A., & Rouet, J. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior, 24,* 753-765.

Leeden R., Meijer E., & Busing F.M. (2008). Resampling multilevel models. In: Leeuw J.., Meijer E. (Eds), *Handbook of Multilevel Analysis.* Springer, New York, NY.

Leite, W., Sandbach, R., Jin, R., MacInnes, J., & Jackman, M. (2012). An evaluation of latent growth models for propensity score matched groups. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(3), 437-456.

Leite, W., Jimenez, F., Kaya, Y., Stapleton, L., MacInnes, J., & Sandbach R. (2015) An evaluation of weighting methods based on propensity scores to reduce selection bias in multilevel observational studies, *Multivariate Behavioral Research, 50*(3), 265-284.

Leite, W. (2017). *Practical Propensity Score Methods Using R.* Sage Publications Ltd.

Leite, W., Aydin, B., & Gurel, S. (2019). A comparison of propensity score weighting methods for evaluating the effects of programs with multiple versions. *Journal of Experimental Education, 87*(1), 75-88.

Levy, F., & Murnane, R. J. (2004). *The New Division of Labor: How Computers Are Creating the Next Job Market.* Princeton, NJ: Princeton University Press.

Lennon, M., Kirsch, I., Von Davier, M., Wagner, M., & Yamamoto, K. (2003). *Feasibility Study for the PISA ICT Literacy Assessment.* Princeton, NJ: Educational Testing Services.

Lohr, S. (1999). *Sampling: Design and Analysis.* Duxbury Press, Pacific Grove, CA.

Lunceford, J., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine, 23*(19), 2937–2960.

Martens, E., Pestman, W., de Boer, A., Belitser, S., & Kungel, O. (2006) Instrumental variables: applications and limitations. *Epidemiology, 17*(3).

Massing, N., & Schneider, S. (2017). Degrees of competency: The relationship between educational qualifications and adult skills across countries. *Large-scale Assessments in Education, 5*(6).

McCaffrey, D., Griffin, B., Almirall, D., Slaughter, M., Ramchand, R., & Burgette, L. (2014). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Stat Med, 32*(19), 3388-3414.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econooometrics.* Academic Press.

Morgan, S., & Harding, D. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research, 35,* 3-60.

Morgan, S., & Winship, C. (2007). *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* New York: Cambridge University Press.

Murnane, R., & Willett, J. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford: Oxford University Press.

National Center for Education Statistics. (2010). *An introduction to NAEP.* U.S. Department of Education. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/parents/2010468.pdf

National Center for Education Statistics. (2013). *The nation's report card: Trends in academic progress 2012 (NCES 2013-456).* Washington, D.C. Retrieved from https://static1.squarespace.com/static/51bb74b8e4b0139570ddf020/t/588a18c13a04118a0c68116e/1485445313820/Arora_Pawlowski_PIAAC_2017.pdf

National Center for Education Statistics. (2017). *Collaborative problem solving: Considerations for the National Assessment of Educational Progress.* U.S. Department of Education. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/researchcenter/collaborative_problem_solving.pdf

National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, D.C.: The National Academies Press. Retrieved from https://doi.org/10.17226/13398.

National Association of Colleges and Employers. (2010). *NACE research report: Job outlook 2010.* Retrieved from https://www.naceweb.org/

Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691-692.

Newell, A., & Simon, H.A. (1972). *Human Problem Solving.* Englewood Cliffts, NJ: Prenctice Hall.

Nicholas, J., & Gulliford, M. (2008). Commentary: What is a propensity score? *The British Journal of General Practice, 58*(555), 687.

OECD & Statistics Canada. (2000). *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*. Paris: OECD.

OECD & Statistics Canada. (2005). *Learning a Living: First Results of the Adult Literacy and Life Skills Survey*. Paris: OECD.

OECD (2001). *Thematic Review on Adult Learning.* Paris: OECD.

OECD (2003). *Problem Solving for Tomorrow's World: First Measures of Cross-curricular Competencies from PISA 2003.* Paris: OECD.

OECD (2009). *PIAAC Background Questionnaire JRA V5.0: Conceptual Framework.* OECD Programme for the International Assessment of Adult Competencies. Retrieved from http://www.oecd.org/edu/48865373.pdf

OECD. (2010). *The OECD Programme for the International Assessment of Adult Competencies (PIAAC).* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/edu/skills-beyond-school/41690983.pdf

OECD. (2011). *PIAAC Conceptual Framework of the Background Questionnaire.* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/PIAAC(2011_11)MS_BQ_ConceptualFramework_1%20Dec%202011.pdf

OECD. (2012). *Literacy, Numeracy and Problem Solving in Technology-Rich Environments: Framework for the OECD Survey of Adult Skills.* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/PIAAC%20Framework%202012--%20Revised%2028oct2013_ebook.pdf

OECD. (2013a). *Technical Report of the Survey of Adult Skills (PIAAC).* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

OECD. (2013b). *The Survey of Adult Skills: Reader's Companion.* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/Skills%20(vol%202)-Reader%20companion--v7%20eBook%20(Press%20quality)-29%20oct%200213.pdf

OECD. (2013c). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills.* Paris, France: OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/Skills%20volume%201%20(eng)--full%20v12--eBook%20(04%2011%202013).pdf

OECD. (2013d). *Skilled for Life? Key Findings from the Survey of Adult Skills.* Retrieved from http://www.oecd.org/skills/piaac/SkillsOutlook_2013_ebook.pdf

OECD. (2013e). *Time for the U.S. to Reskill? What the Survey of Adult Skills Says.* Paris, France: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/docserver/download/8713041e.pdf?expires=1501257889&id=id&accname=guest&checksum=6FA9F09FA3E2C51D872D01F6341EF630

OECD. (2015). *PISA 2015 Results: Collaborative Problem Solving. Volume V.* Retrieved from http://www.izm.gov.lv/images/aktualitates/2017/Ptjuma-rezultaati.pdf

OECD. (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. OECD Skills Studies, Paris, France: OECD Publishing, Paris.

OECD. (2017). *PISA 2015 Collaborative Problem-Solving Framework.* Retrieved from https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Paccagnella, M. (2016). *Literacy and Numeracy proficiency in IALS, ALL and PIAAC.* (OECD Education Working Papers, No. 142). OECD Publishing, Paris. Retrieved from http://dx.doi.org/10.1787/5jlpq7qglx5g-en

Partnership for the 21st Century Skills. (2010). *21st Century Readiness for Every Student: A Policymaker's Guide.* Tucson, AZ: Author. Retrieved from http://www.p21.org/documents/policymakersguide_final.pdf

Petryna, A. (2009). *When Experiments Travel: Clinical trials and the global search for human subjects.* Princeton, NJ: Princeton University Press.

PIAAC Literacy Expert Group (2009a). *PIAAC Literacy: A Conceptual Framework* (OECD Education Working Paper No. 34). Paris, France: OECD Publishing.

PIAAC Numeracy Expert Group (2009b). *PIAAC Numeracy: A Conceptual Framework* (OECD Education Working Paper No. 35). Paris, France: OECD Publishing.

PIAAC Expert Group in Problem Solving in Technology-Rich Environments (2009c). *PIAAC Problem Solving in Technology-Rich Environments: A Conceptual Framework* (OECD Education Working Paper No. 36). Paris, France: OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/docserver/download/220262483674.pdf?expires=1505873612&id=id&accname=guest&checksum=35A446C07A20A9CD6945FDB160D27C76

Podolsky, S., Jones, D., & Kaptchuk, T. (2016). From trials to trials: Blinding, medicine, and honest adjudication. In: Robertson, C. & Kesselheim, A. (Eds.), *Blinding as a Solution to Bias: Strengthening Biomedical Science, Forensic Science, and Law.* London: Elsevier.

Rassen, J., Glynn, R., Brookhart, M., & Schneeweiss, S. (2011). Covariate selection in high-dimensional propensity score analyses of treatment effects in small samples. *Am. J. Epidemiol, 173,* 1404–1413.

Reardon, S., & Raudenbush, S. (2009). Assumptions of value□added models for estimating school effects. *Education Finance and Policy, 4*(4), 492-519.

Reder, S. (2009a). The development of adult literacy and numeracy in adult life. In: Reder, S., & Bynner, J. (Eds.), *Tracking Adult Literacy and Numeracy Skills: Findings from Longitudinal Research.* Routledge, New York, 59-84.

Reder, S. (2009b). Scaling up and moving. In: Connecting social practices views to policies and programs in adult education. *Literacy and Numeracy Studies, 16*(2), 35-50.

Resnik, D. (2008). Randomized controlled trials in environmental health research: Ethical issues. *J Environ Health, 70*(6), 28-30.

Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. (2017). *Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the twang package.* Retrieved from https://cran.r-project.org/web/packages/twang/vignettes/twang.pdf

Robins, J., Hernan, M., & Brumback, B. (2000). Structural models and causal inference in epidemiology. *Epidemiology*, *11*(5), 550-560.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical*

Association, 79, 516-524.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician, 39,* 33–38.

Rosenbaum P. (1987). Optimal matching for observational studies. *Journal of the American Statistical Association, 84*(408), 1024.

Rosenbaum P. (2002). *Observational Studies.* New York, NY: Springer.

Rosenbaum P. (2005). Observational study. *Encyclopedia of Statistics in Behavioral Science, 3,* 1451-1462.

Rubin, D. (1973a). Matching to remove bias in observational studies, *Biometrics, 29,* 159-183.

Rubin, D. (1973b). The use of matched sampling and regression adjustments to remove bias in observational studies, *Biometrics, 29,* 185-203.

Rubin, D. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. *Journal of Education Psychology, 66,* 688–701.

Rubin, D. (1986). Comment: Which ifs have causal answers? *Journal of the American Statistical Association, 81,* 961-962.

Rubin, D. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127,* 757-763.

Schafer, J., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*(4), 279-313.

Schneeweiss, S., Rassen, J., Glynn, R., Avorn, J., Mogun, H., & Brookhart, M. (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology* 20, 512.

Schwab, J. (2002). *Multinomial Logistic regression: Basic relationships and complete problems.* Retrieved from http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/

Sgobbi, F. (2014). *Adult education, training and individual performance: Some preliminary evidence from PIAA.* Paper presented at the XXIX National Conference of Labor Economics, University of Pisa, Pisa.

Shadish, W. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15*(1), 3-17.

Starkweather, J., & Moske, A. (2011). *Multinomial logistic regression.* Retrieved from https://it.unt.edu/sites/default/files/mlr_jds_aug2011.pdf

Stuart, E., & Rubin, D. (2007). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. Osboarne (Ed.), *Best Practices in Quantitative Methods* (pp. 155-176). Thousand Oaks, CA: Sage.

Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science,* 25(1), 1-21.

Thoemmes, F., & Kim, E.S. (2011). A systematic review of propensity score methods in the social sciences. *Multivariate Behavioral Research, 46*(1), 90-118.

Torgerson C., Brooks G., Porthouse J., Burton M., Robinson A., Wright K., & Watt I. (2004). *Adult literacy and numeracy interventions and outcomes: A review of controlled trials*. London: National Research and Development Centre for Adult Literacy and Numeracy.

Tukey, J. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics, 5*(2), 99–114.

UNESCO. (2011), *Revision of the International Standard Classification of Education (ISCED).* Paper 36 C/19, 34th session of the General Conference, 2011, UNESCO.

UNESCO Institute for Statistics. (2012). *International Standard Classification of EDucation ISCED 2011*. Retrieved from http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf

von Davier, M., Gonzalez, E., & Mislevy, R. (2009) What are plausible values and why are they useful? In von Davier, M., & Hastedt, D. (Eds.), IERI *Monograph Series: Issues and Methodologies in Large Scale Assessments: Volume 3.* IERInstitute.

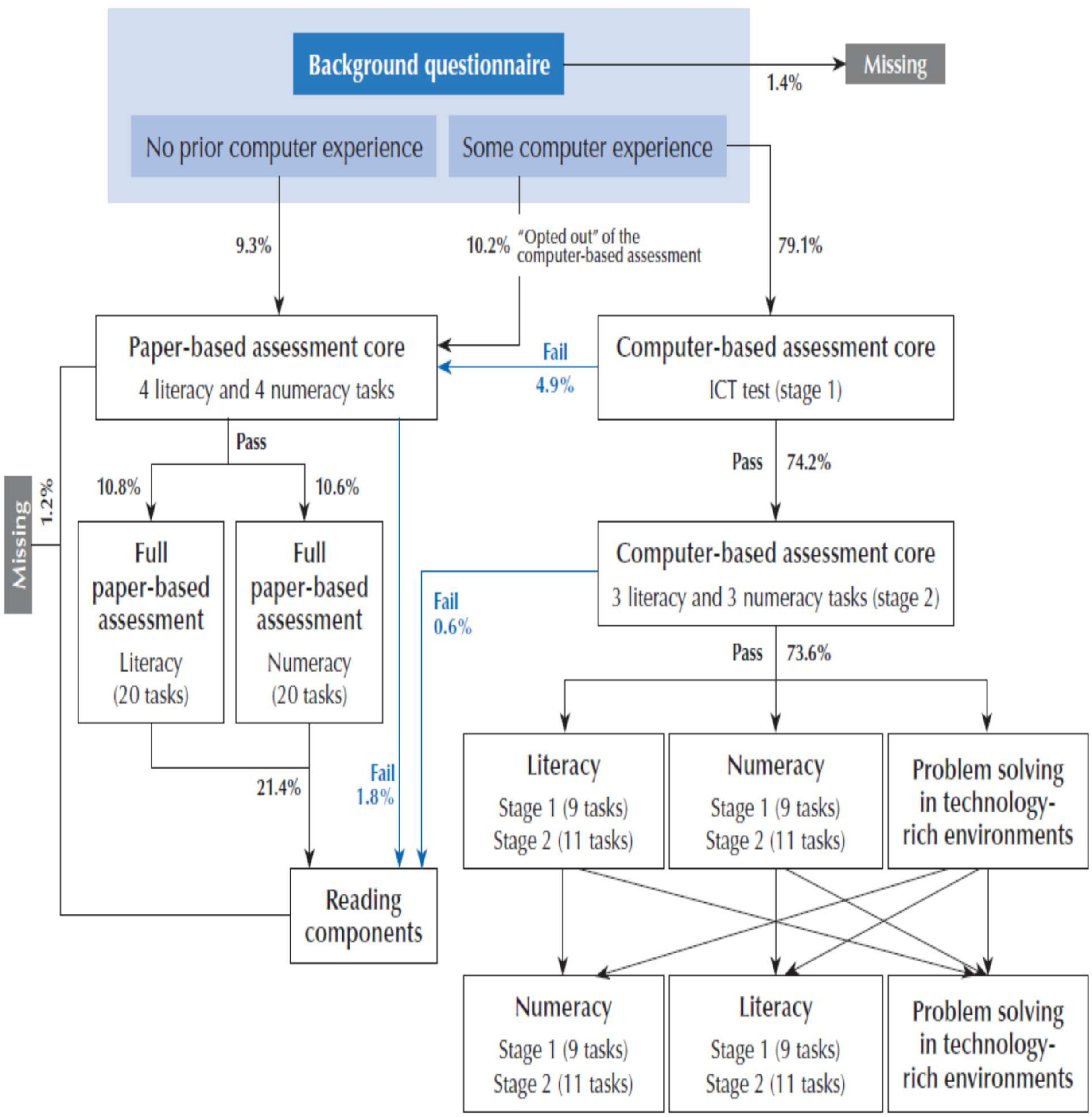
Appendix A. Participating Countries in PIAAC (round 1) and Their Sample Sizes

National entities	Cognitive domains assessed	Assessment language(s)	Groups oversampled	Achieved sample
Australia	L, N, PS-TRE	English	Persons resident in certain states and territories	7 428
Austria	L, N, PS-TRE	German		5 130
Canada	L, N, PS-TRE	English, French	Persons aged 16-25, provinces/territories, linguistic minorities, aboriginal persons, and recent immigrants	27 285
Czech Republic	L, N, PS-TRE	Czech	Persons aged 16-29	6 102
Denmark	L, N, PS-TRE	Danish	Persons aged 55-65 years, recent immigrants	7 328
Estonia	L, N, PS-TRE	Estonian, Russian		7 632
Finland	L, N, PS-TRE	Finnish, Swedish		5 464
France	L, N	French		
Germany	L, N, PS-TRE	German		5 465
Ireland	L, N, PS-TRE	English		5 983
Italy	L, N	Italian		4 621
Japan	L, N, PS-TRE	Japanese		5 278
Korea	L, N, PS-TRE	Korean		6 667
Netherlands	L, N, PS-TRE	Dutch		5 170
Norway	L, N, PS-TRE	Norwegian		5 128
Poland	L, N, PS-TRE	Polish	Persons aged 19-26	9 366
Slovak Republic	L, N, PS-TRE	Slovak, Hungarian		5 723
Spain	L, N	Castilian, Basque, Catalan, Galician, Valencian		6 055
Sweden	L, N, PS-TRE	Swedish		4 469
United States	L, N, PS-TRE	English		5 010
Sub-national entities				
Flanders (Belgium)	L, N, PS-TRE	Dutch		5 463
England (UK)	L, N, PS-TRE	English		5 131
Northern Ireland (UK)	L, N, PS-TRE	English		3 761
Partner				
Cyprus[1]	L, N	Greek		5 053

1. See notes at the end of this chapter.
Note: L = Literacy, N = Numeracy and PS-TRE = Problem Solving in Technology-Rich Environments.

Note. From *The Survey of Adult Skills: Reader's companion,* p. 54, by OECD, 2013. Retrieved from http://www.oecd.org/skills/piaac/Skills%20(vol%202)-Reader%20companion--v7%20eBook%20(Press%20quality)-29%20oct%200213.pdf

Appendix B. Percentages of Respondents Taking Different Pathways in PIAAC



Note: The figures presented in this diagram are based on the average of OECD countries participating in the Survey of Adult Skills (PIAAC).

Note. From *The Survey of Adult Skills: Reader's companion,* p. 49, by OECD, 2013. Retrieved from http://www.oecd.org/skills/piaac/Skills%20(vol%202)-Reader%20companion--v7%20eBook%20(Press%20quality)-29%20oct%200213.pdf

Appendix C. Indicators of Skills Use at Work and Sample Tasks

	Indicator	Group of tasks
Information-processing skills	Reading	Reading documents (directions, instructions, letters, memos, e-mails, articles, books, manuals, bills, invoices, diagrams, maps)
	Writing	Writing documents (letters, memos, e-mails, articles, reports, forms)
	Numeracy	Calculating prices, costs or budgets; use of fractions, decimals or percentages; use of calculators; preparing graphs or tables; algebra or formulas; use of advanced math or statistics (calculus, trigonometry, regressions)
	ICT skills	Using e-mail, Internet, spreadsheets, word processors, programming languages; conducting transactions on line; participating in online discussions (conferences, chats)
	Problem solving	Facing complex problems (at least 30 minutes of thinking to find a solution)
Other generic skills	Task discretion	Choosing or changing the sequence of job tasks, the speed of work, working hours; choosing how to do the job
	Learning at work	Learning new things from supervisors or co-workers; learning-by-doing; keeping up-to-date with new products or services
	Influencing skills	Instructing, teaching or training people; making speeches or presentations; selling products or services; advising people; planning others' activities; persuading or influencing others; negotiating.
	Co-operative skills	Co-operating or collaborating with co-workers
	Self-organising skills	Organising one's time
	Dexterity	Using skill or accuracy with one's hands or fingers
	Physical skills (gross)	Working physically for a long period

Note. From *OECD skills outlook 2013: First results from the Survey of Adult Skills,* p. 143, by OECD, 2013, OECD Publishing. Retrieved from http://www.oecd.org/skills/piaac/Skills%20volume%201%20(eng)--full%20v12--eBook%20(04%2011%202013).pdf

Appendix D. Accounting for imputation error variance component

For estimation using plausible values (PVs), calculations must account for both the sampling error component and the variance due to imputation of proficiency scores. The estimator of the population mean is the average of the M PV means,

$$\hat{\bar{Y}}^* = \sum_{m=1}^{M} \hat{\bar{Y}}_m \Big/ M.$$

The variance of the estimated mean $\hat{\bar{Y}}^*$ is computed using formulas specific to PVs as follows:

$$v\left(\hat{\bar{Y}}^*\right) = U^* + B\left(1 + \frac{1}{M}\right)$$

where, the "within" variance component is computed as the average of the sampling variance for each of the M plausible values, computed as,

$$U^* = \left(\sum_{m=1}^{M} U_m\right) \Big/ M$$

where the sampling variance of the estimated mean $\hat{\bar{Y}}_m$ for plausible value m is U_m, and

where, the "between" component is calculated as

$$B = \left[\sum_{m=1}^{M} \left(\hat{\bar{Y}}_m - \hat{\bar{Y}}^*\right)^2\right] \Big/ (M-1)$$

where, the mean of each of the M PVs $y_{l1}, y_{l2}, \ldots, y_{lM}$ for sample unit l is computed as

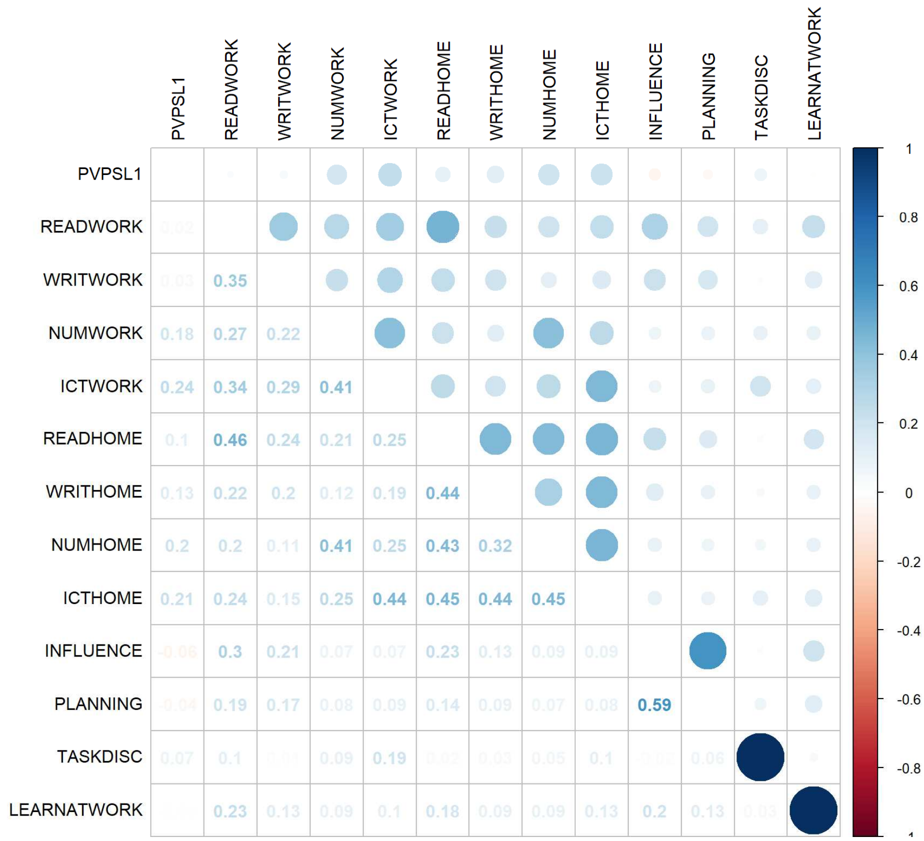
$$\hat{\bar{Y}}_m = \sum_{l \in s} w_l y_{lm} \Big/ \sum_{l \in s} w_l \; ; m = 1, \ldots, M,$$

where s denotes the set of sample units.

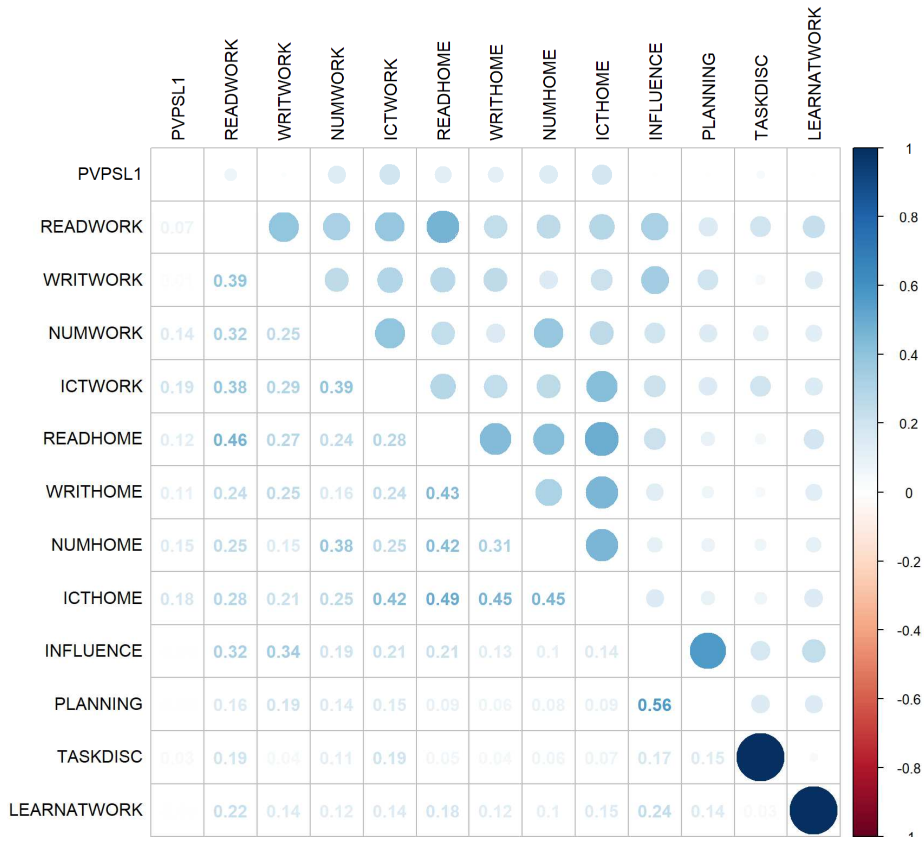
The standard error is computed as the square root of the total variance, $\sqrt{v\left(\hat{\bar{Y}}^*\right)}$.

Note. From *Technical report of the Survey of Adult Skills (PIAAC),* ch.15, p.29, by OECD, 2013. Retrieved from http://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

Appendix E. Correlation Matrix between Skills Used at Work for Full-time Professionals



Appendix F. Correlation Matrix between Skills Used at Work for Full-time Associates



Appendix G. Results from One-Way ANOVA Post-Hoc Tests

There are significant between-country variances in correlations (between PSL & NUM/LIT).

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## cor_PN$CNTRYID 13  0.694  0.0534    2.02  0.027 *
## Residuals      98  2.595  0.0265
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##                 Df Sum Sq Mean Sq F value Pr(>F)
## cor_PL$CNTRYID 13   0.86  0.0662    2.23  0.014 *
## Residuals      98   2.91  0.0297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## cor_NL$CNTRYID 13   1.37  0.1054    4.47 6.4e-06 ***
## Residuals      98   2.31  0.0236
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```