A High-Stakes Approach to Response Time Effort in Low-Stakes Assessment

Munevver Ilgun Dibek[*] (iD)
TED University, TURKEY

Abstract: Response times are one of the important sources that provide information about the performance of individuals during a test process. The main purpose of this study is to show that survival models can be used in educational data. Accordingly, data sets of items measuring literacy, numeracy and problem-solving skills of the countries participating in Round 3 of the Programme for the International Assessment of Adult Competencies were used. Accelerated failure time models have been analyzed for each country and domain. As a result of the analysis of the models in which various covariates are included as independent variables, and response time for giving correct answers is included as a dependent variable, it was found the associations between the covariates and response time for giving correct answers were concluded to vary from one domain to another or from one country to another. The results obtained from the present study have provided the educational stakeholders and practitioners with valuable information.

Keywords: *Accelerated failure time model, PIAAC, response time, survival analysis.*

Introduction

When the test is a low-stakes test for respondents, it is likely that the respondents tend not to perform at their best (Finn, 2015). In such a situation, respondents do not focus on the extent to which their scores on these tests reflect their true ability level, because these results do not have a significant impact on their lives (Wise & DeMars, 2005). Therefore, it is important for test givers to obtain evidence of the validity for the results obtained from such tests. Additionally, there may be the respondents who perform at their best although the nature of the tests they take is low stakes. Thus, it is difficult to differentiate how many of the respondents actually performed at their best and how many did not spend enough time answering the test items. Accordingly, it is also difficult to determine the extent to which low-test effort causes biases in the results obtained from the tests (Wise & Kong, 2005). All in all, because results from low-stake assessments show "what students will demonstrate with minimal effort" (O'Neil et al., 1995, p.135) rather than what they know, investigating the efforts respondents make in answering test items is critical for policy makers and educators to make valid interpretations based on test results.

Focusing on the time between the moment the item is presented to the respondents and the moment the item is answered, namely the response time (RT) is a more effective method in determining the respondents' efforts in the test. To give further explanation, the respondents who take the test in computer environment are not aware that their reaction times are automatically recorded, which does not affect their effort to respond, thus avoiding bias. Therefore, this study focuses on the RTs of the respondents participating in the Programme for the International Assessment of Adult Competencies (PIAAC) 2017 as their interactions with the computer are automatically logged while answering the items. The reason why PIAAC is chosen among other low-stake assessments is that it measures the information-processing skills of adults such as literacy, numeracy and problem solving in technology-rich environments, which ensure full participation in today's society where 21st century skills are important and necessary for the labor market, education and training, (OECD, 2013a). Therefore, the frequent use of these skills in many points of our lives makes it necessary to focus on these skills and RTs of adults for answering the items requiring the use of these skills in the present study.

---

[*] **Correspondence:**
Munevver Ilgun Dibek. TED University, Department of Educational Sciences, Ziya Gökalp Street, Ankara, Turkey. ✉ munevver.ilgun@tedu.edu.tr

Models for Response Times and Process in Testing

There are various types of latent trait models for response time modelling in psychometrics and educational measurement. Approaches that model response time can be grouped under four headings: response time models, joint models, local dependence models and response times as covariate models (De Boeck & Jeon, 2019). In response time models, response time is considered as a dependent variable. These models are divided into three sub-models: (1) distribution models for response time (e.g., Ranger & Ortner, 2012; van Zandt, 2002), (2) explanatory models (e.g., Sternberg, 1985) and (3) models with response accuracy as a covariate (e.g., Novikov et al., 2017). For example, in their study, Ranger and Ortner (2012) proposed to use Cox model in RT modelling, and they replaced the observed covariates in the Cox model with the test takers' latent speed. In joint models, response time together with other variable (e.g., accuracy) are handled as dependent variable (De Boeck & Jeon, 2019). In the literature, joint models are divided into three as hierarchical models (e.g., van der Linden, 2007; Wang et al., 2013), diffusion model (van der Maas et al., 2011) and race model (Rouder et al., 2015). For instance, Wang et al. (2013) modelled RT and response accuracy simultaneously. In local dependence models, response time and other variable (e.g., response accuracy) are jointly modelled with the possibility of dependencies beyond dependencies captured by latent variables and item parameters (De Boeck & Jeon, 2019). On the other hand, response times as covariate models include response times as an independent variable. These models are divided into two as SAT-Based Models (Heitz, 2014) and GLMM Based Covariate Models (e.g., Goldhammer et al., 2014). For more general review of subtypes of these models, see review by De Boeck and Jeon (2019). Most recently, it has been seen that there are several studies linking various personal characteristics with the RTs of adults in educational testing (e.g., Wang & Chen, 2020). In their study, Wang and Chen (2020) used response times and response accuracy as test takers' latent constructs to measure their fluency within cognitive diagnosis models. In this respect, according to the classification made by De Boeck and Jeon (2019), the model used by the researchers can be described as the response times as covariate model. Additionally, Ranger et al. (2020) extended van der Linden (2007)'s the hierarchical model that allows for response times shifts via an additional location parameter.

Among the models classified by De Boeck and Jeon (2019), survival models are one of the models placed under the heading of distribution models for response time. Survival models for response time can be considered as a high stakes approach. More precisely, this high stakes nature of parametric survival models comes from the fact that they are easier to implement and interpret compared to many models suggested in the literature (Swindell, 2009). Specifically, under accelerated failure time (AFT) models, which is one of the parametric survival models, one can determine the direct impact of the explanatory variables on the survival time instead of hazard, as it is done in the proportional hazards (PH) model. This characteristic allows for an easier interpretation of the results because the parameters measure the impact of the correspondent covariate on the mean survival time (Bekalo, 2019).

In this context, the outline of this manuscript is as follows. Firstly, the models used in survival analysis are described. Secondly, its application to RTs in cognitive tests used in PIAAC 2017 as a large-scale assessment is demonstrated since differential test efforts and RTs confound the test scores of different countries participating in international studies (Wise & DeMars, 2010). Furthermore, several covariates for estimating the RTs were included to better understand the adults' testing efforts.

Survival Analysis

Survival analysis is a set of methods used in examining and modelling the time that passes until the occurrence of a defined event. Survival analysis is widely used in the literature, especially in epidemiology and medical studies. In addition, its use has been increasing especially in the field of education in recent years. "If viewing 'giving a response' as an event, RT shares the same meaning as the survival time in biostatistics, and therefore RTs can be modelled similarly" (Wang et al., 2013, p.383). Survival analysis presents a standard approach regarding the time that passes until a specific event occurs, and it has no other assumptions except that the hazard ratio is constant versus time. It also includes outliers in the analysis as a censored observation.

In survival analysis, the occurrence of the event of interest is defined as a failure (Chernick & Friis, 2003). The time between a specific start time of a unit and the occurrence of the event of interest is called the failure time and is usually denoted by T. In survival analysis, the event is usually observed for some units in the events, while the time of the event is greater than the current follow-up time for other events and the event may not be observed. These types of events are frequently encountered in survival analysis and in this case, the event is called a censored event. The main advantage of survival analysis compared to classical statistical methods is that censored events are also included in the analysis (Kleinbaum & Klein, 2005). Survival analysis mainly deals with the risk or hazard of failure that occurs at any time from the starting point of the study.

Models that investigate the effects of covariates on survival time as a dependent variable have an important place in survival analysis. In survival analysis, these models are divided into three groups as non-parametric, semi-parametric and parametric models. Kaplan Meier (KM) estimate is one of the most used nonparametric methods to predict the survival function. The Cox regression model (proportional hazard model-CRM) proposed by Cox (1972) is a semi

parametric regression model popularly used in survival analysis (Fox & Weisberg, 2011). This model examines the relationship between survival time and one or more covariates (Singh & Mukhopadhyay, 2011). Although semi-parametric survival models have the advantage of flexibility, they may suffer from reduced conciseness when compared to parametric survival models. More precisely, once the appropriate statistical distribution is chosen, a parametric survival model will provide somewhat greater efficiency because it estimates fewer parameters (Wheatley-Price et al., 2012). Additionally, in order to use the CRM, the proportional hazard assumption must be provided. When the proportional hazard assumption is not provided, it is appropriate to use either CRMs with stratified or time-dependent covariates or parametric regression models (parametric proportional hazard models or accelerated failure time models) instead of the simple CRM (Kleinbaum & Klein, 2005).

Parametric Survival Model

Parametric regression models have more effective regarding parameter estimates than CRMs (Nardi & Schemper, 2003). Parametric models for survival data can be used as parametric proportional hazard (PH) models and accelerated failure time (AFT) models. PH models are a parametric modification of the CRM and they assume that the survival time has a certain distribution. However, because the number of the probability distributions that can be used with PH models is limited compared to those that can be used with AFT models, AFT models are more commonly preferred in survival analysis. 'AFT model has been fitted for distributions like exponential, weibull, log-normal, log-logistic, gamma etc.' (Saikia & Barman, 2017, p.320). Accordingly, AFT models are named according to the name of the distribution to which they fit.

The AFT model assumes that the covariates are linear with the logarithm of the survival time T and is expressed as,

$$logT_i = \mu + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \sigma\epsilon_i \qquad (1)$$

where μ and α are regression coefficients, σ is the scale parameter, X_p denotes covariates and ϵ_i is the error term.

The covariates are often included in the model through the hazard function $h(t)$. The hazard function is the instantaneous rate at which events occur. In psychological terms, the hazard rate is the conditional probability of finishing the task in the next moment. In other words, the hazard rate measures the individual's relative ability to perform mental work in a unit of time. High hazard rates indicate that the examinee works more intensely (Wenger & Gibson, 2004). The survival function S(t) and hazard function $h(t)$ of these models are given in Table 1 (Qui, 2009; Wei, 1992).

Table 1. The survival and hazard functions for AFT models

	$S(t)$	$h(t)$
Weibull	$\exp\left[-\exp\left(\frac{-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-\cdots-\alpha_p x_{pi}}{\sigma}\right)t^{\frac{1}{\sigma}}\right]$	$\frac{1}{\sigma}t^{\frac{1}{\sigma}-1}\exp\left(\frac{-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-\cdots-\alpha_p x_{pi}}{\sigma}\right)$
Log-Normal	$1-\phi\left(\frac{\log(t)-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-\cdots-\alpha_p x_{pi}}{\sigma}\right)$	$\frac{\frac{1}{t\sigma\sqrt{2\pi}}\left[-\frac{1}{2\sigma^2}(\log(t)-\mu)^2\right]}{1-\phi\left(\frac{\log(t)-\mu}{\sigma}\right)}$
Log-Logistic	$\left[1+\exp\left(\frac{-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-\cdots-\alpha_p x_{pi}}{\sigma}\right)t^{\frac{1}{\sigma}}\right]^{-1}$	$\frac{1}{\sigma t}\left[1+\exp\left(\frac{-\mu-\alpha_1 x_{1i}-\alpha_2 x_{2i}-\cdots-\alpha_p x_{pi}}{\sigma}\right)t^{\frac{1}{\sigma}}\right]^{-1}$

Note. ϕ is cumulative density function of standard normal distribution.

The maximum likelihood estimation method is used to obtain AFT model parameters. The likelihood function for n observed survival functions $t_1, t_2, \ldots, t_n$ can be written as,

$$L = \prod_{i=1}^{n}\{f_i(t_i)\}^{\delta_i}\{S_i(t_i)\}^{1-\delta_i} \qquad (2)$$

where $S_i(t_i)$ and $f_i(t_i)$ represent the survival and probability density functions for the i-th unit at t_i, respectively, and δ_i is called the event indicator for the i-th unit. The event indicator represents the censored event when $\delta_i = 0$ and failure when $\delta_i = 1$.

The log likelihood function with $S_i(t_i) = S_{\epsilon_i}(z_i)$ can be obtained as,

$$logL = \sum_{i=1}^{n}\sigma t_i^{-\delta_i}\left\{f_{\epsilon_i}(z_i)\right\}^{\delta_i}\left\{S_{\epsilon_i}(z_i)\right\}^{1-\delta_i}$$

$$logL = \sum_{i=1}^{n}-\sigma_i \log(\sigma t_i) + \delta_i \, log\{f_{\epsilon_i}(z_i)\} + (1-\delta_i)\{S_{\epsilon_i}(z_i)\} \qquad (3)$$

where $z_i = \frac{logt-\mu-\beta_1 x_1-\cdots-\beta_p x_p}{\sigma}$. To find the unknown parameters μ, σ and $\beta_1, \ldots, \beta_p$, the log likelihood function given in equation (3) should be maximized by using the Newton Raphson procedure (Saikia & Barman, 2017).

The covariates for which their associations with the response time spent to give correct answers are explained in the following part.

Covariates of Response Time

One of the biggest difficulties in measuring the true RTs of respondents is that they are affected by various factors. Specifically, several individual-level characteristics can influence the RTs of individuals in testing. In this context, in the computer based assessment part of PIAAC 2017, examining the variables that affect RTs is crucial to make valid interpretations regarding the proficiency of the respondents. In the related literature focusing on survival analysis, predictor variables are usually termed covariates. Therefore, the following covariates are addressed in the present study.

The number of actions

With the implementation of the PIAAC assessment in computer environment, in addition to the RTs, data on the number of actions of the respondents such as clicking a button or writing an answer before answering each item can also be collected (OECD, 2019a). Although this information provides important information about the performances of the respondents and reduces measurement errors in large-scale assessments, the relationship between the factors regarding the process such as the number of actions and the competencies and test-efforts of the respondents has not been sufficiently revealed in the literature. However, the high number of actions can be associated with the respondents not having enough non-cognitive skills, level of self-confidence or hesitative behaviors (Anghel & Balart, 2017). Therefore, in order to better understand the adults' response process, the number of actions of adults is also addressed in the present study.

Gender

The performance of respondents of different gender in the test and their efforts in responding to the test items differ, and this differentiation may vary according to the feature to be measured by the test. For example, females can achieve higher success in verbal tests that require the use of reading skills, while males can achieve higher success in mathematics and science tests that require the use of numerical skills (Quinn & Cooc, 2015). This difference in achievements stemming from gender is also seen in large-scale assessments (Balart & Oosterveen, 2019). Therefore, in order to determine how this situation occurs in adult groups, the effect of the gender has been investigated in the present study.

Educational attainment

It is not possible to directly say that educational attainment levels, which mean the level of formal education completed by individuals (OECD, 2020), have a positive or negative relationship with their performance on literacy, numeracy and problem solving, which are some of the points examined in this study. Specifically, some studies have linked not having sufficient skills in these areas with being low-educated, dropping out of school at an early age, being in a socially deprived community, and working in simple jobs (e.g., Kureková et al., 2013). Additionally, although adults have developed sufficient literacy, numeracy or problem solving skills in their formal education, they may have lost these skills when they have not used them later (OECD, 2013b). Therefore, examining the education levels of adults is valuable in understanding their test-taking efforts.

Age

It is important to examine the effects of age on the performance of individuals in a certain area in the assessment practices involving respondents with ages ranging from young to old. Especially, when these assessments are made in computer environment, the difference in the performance of the individuals may be more pronounced due to their ages. More precisely, although older adults historically have little use of computers, the frequency of computer use of this population consisting of baby-boomers (i.e., adults born between the years of 1946 -1964) is increasing rapidly (Hart et al., 2008). Although the frequency of using computers among older adults has increased, studies in the literature show that compared to younger individuals, their level of familiarity with computers and exposure to computers are still lower (Harvey et al., 2020). This situation results in a low level of performance. In addition, depending on their age, they tend to respond more slowly to the items and have longer non-decision time due to cognitive aging (Ratcliff et al., 2004). Therefore, in order to determine how this situation happens in a low-stakes assessment, the effect of the age has been investigated in the present study.

Readiness to learn

Readiness to learn is a complex structure that includes attitudinal, emotional (pleasure), cognitive (metacognitive skills), behavioral (time management) and personal components. Therefore, it is not possible to say that there is a simple relationship between such a complex structure and the performance of the individuals in the test. Individuals with a high level of readiness to learn are eager to learn the information presented to them, exhibit a high level of engagement, are determined to and concentrated on completing the learning tasks presented to them (Eccles & Wigfield, 2002). Some researches (e.g, Reder, 1998; Smith et al., 2015) have pointed out the mediating effect of

individuals' use of readiness to learn strategies on individuals' competency levels in the test. For example, levels of individuals' readiness to learn play a role as a mediator variable between educational attainments and ability levels. To give further explanation, a high level of readiness to learn reduces the differences between the competency levels of individuals with high and low educational attainment levels (Smith et al., 2015). As a result, in this study, readiness to learn is handled as a covariate in order to explain the complex relationship between the levels of individuals' readiness to learn and their performance during the test process in terms of individuals' response times.

Use of Information and Communication Technologies (ICT) Skills

With the advancements in technology, the frequent use of computer-based testing (CBT) made it necessary to look for some features in those who participated in the test. For example, familiarity with ICT tools and their applications on a certain level is presented as a requirement to participate in computerized forms of large-scale assessments such as PIAAC (OECD, 2019b). Additionally, the need arose to examine the factors that have a critical impact on the performance of individuals who have taken tests of large-scale assessments in computer environment. Several studies (e.g., Sawaki, 2001) state that individuals' familiarity with computers do not have a significant effect on their performance, while some studies (e.g, Odo, 2012) state that this effect may change depending on how similar the test items presented to them in computer environment are to those presented to them in paper-pencil tests. For example, individuals who have the habit of answering the questions by highlighting them in paper-pencil tests may have the tendency to get lower scores if they cannot continue this habit in computer environment (Odo, 2012). In addition, there are studies indicating that the ability to use technological tools related to computers such as mouse or keyboard also affects their speed and success in tests (e.g, Pomplun et al., 2002). Therefore, this inconsistency in research findings appears to warrant further investigation.

Methodology

Population and Sample

The target population of the present study comprised individuals from 16 to 65 years of age who resided in one of the countries that participated in PIAAC 2017 at the time of data collection (OECD, 2019b). The adults taking the computer-based assessment part of the latest round (i.e., Round 3) of the first cycle of PIAAC in countries such as Ecuador, Hungary, Kazakhstan, Mexico, Peru and United States were selected as the sample of this study. The sample was selected from adults aged between 16 and 65 from each country using a multistage sample design for each stage of selection. Specifically, the number of participants in each country is presented in Table 2.

Table 2. The Number of the Participants in PIAAC 2017

Covariate	Subcategory	Mexico (n_1 =2338)	Peru (n_2=3131)	Kazakhstan (n_3 =4412)	Hungary (n_4 =3943)	USA (n_5 =2832)	Ecuador (n_6=2345)
Cognitive pre-test	3	150	319	113	79	87	237
	4	349	574	315	228	225	418
	5	808	1048	1222	848	756	811
	6	1031	1190	2762	2788	1764	879
Gender	1=Male	1188	1603	1694	1881	1298	1101
	2=Female	1150	1528	2718	2062	1534	1244
Educational Attainment	1=Lower than high school	850	213	455	480	217	651
	2=High school	887	1474	1493	1694	1003	1072
	3=Above high school	599	1442	2459	1764	1606	620
	Missing	2	2	5	5	6	2
Age	1=24 or less	851	998	840	716	435	954
	2=25-34	641	890	1343	831	575	647
	3=35-44	476	642	1075	995	549	433
	4=45-54	254	385	722	758	478	219
	5=55 plus	116	216	432	643	572	92
	Missing		0	0	0	223	0

Table 2. Continued

Covariate	Subcategory	Mexico (n₁ =2338)	Peru (n₂=3131)	Kazakhstan (n₃ =4412)	Hungary (n₄ =3943)	USA (n₅ =2832)	Ecuador (n₆=2345)
Readiness to learn	1=All zero response	3	0	56	4	0	4
	2=Lowest to 20%	131	282	1444	636	179	324
	3=More than 20% to 40%	240	528	771	723	389	343
	4=More than 40% to 60%	298	674	518	693	486	402
	5=More than 60% to 80%	502	874	699	763	616	480
	6=More than 80%	1164	773	922	1124	1162	792
	Missing	0	0	2	0	0	0
Use of ICT skills at home	1=All zero response	22	20	152	14	6	30
	2=Lowest to 20%	333	519	1503	783	234	374
	3=More than 20% to 40%	311	417	760	792	479	246
	4=More than 40% to 60%	273	389	486	703	601	238
	5=More than 60% to 80%	274	424	369	685	585	284
	6=More than 80%	447	751	424	694	643	596
	Missing	678	611	718	272	284	577
Use of ICT skills at work	1=All zero response	36	35	68	110	94	28
	2=Lowest to 20%	200	327	466	404	358	165
	3=More than 20% to 40%	193	304	417	364	347	155
	4=More than 40% to 60%	136	238	302	325	316	119
	5=More than 60% to 80%	163	276	236	428	341	138
	6=More than 80%	298	358	360	556	545	236
	Missing	1312	1593	2563	1756	831	1504

Tools for Data Collection

The data collection of the Survey of Adult Skills was undertaken from July 2017 to the end of December 2017 (OECD, 2019b). The Survey of Adult Skills was administered in two stages: completion of the background questionnaire and completion of the cognitive assessment. Data regarding Round 3 of PIAAC were obtained from the OECD's international PIAAC website (http://vs-web-fs-1.oecd.org/piaac/puf-data/SPSS/).

Background Questionnaire. The background questionnaire (BQ), which was the first part of the PIAAC assessment. It takes up to 30-40 minutes to complete BQ that includes approximately 300 questions measuring the factors which have an effect on the development and maintenance of skills such as education, social background, engagement with literacy and numeracy and ICTs, languages, as well as information on outcomes which may be related to skills (OECD, 2019b).

In this study, several continuous and categorical variables measured with BQ were focused. Specifically, the scores of adults obtained from cognitive pre-test (3, 4, 5, and 6) are taken as continuous variable. Cognitive pre-test includes three literacy and three numeracy items. The remaining variables such as ("male" and "female"), age ("Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more"), educational attainment ("Less than high school", "High school", and "Above high school"), readiness to learn ("All zero response", "Lowest to 20%", "More than 20% to 40%" , "More than 40% to 60%", "More than 60% to 80%", "More than 80%"), the use of ICT skills at home and work of adults ("All zero response", "Lowest to 20%", "More than 20% to 40%" , "More than 40% to 60%", "More than 60% to 80%", "More than 80%") were taken as categorical variables. The categories of the covariate "readiness to learn" derived from six items indicates how intensely the individuals use readiness to learn strategies such as relating new ideas into real life, liking to learn new strategies, attributing something new, getting to the bottom of difficult things, figuring out how different ideas fit together, looking for additional information for clarity (OECD, 2013c). To give as an example, "more than 80%" category of the covariate "readiness to learn" means that readiness to learn strategy use rate of the individuals placed under this category is 80%. In other words, the adults who fall in this highest category are the high readiness to learn strategies users. In the similar manner, the categories of the covariates "ICT skills at home" derived

from seven items and "ICT skills at work" derived from six items. They show how much they use ICT skills at home and work such as the use of computer, e-mail, internet for information to conduct monetary transaction, spreadsheets, word processing and overall level of computer use in terms of complexity (OECD, 2013c). The interpretation of the categories of these covariates are the same with the covariate "readiness to learn".

Cognitive Assessment Tools

Depending on the respondents' initial familiarity or skill with ICT which is assessed by the background questionnaire, they are routed either to a printed assessment (i.e., 'paper-and-pencil'), or to a computer-based assessment. To put in different words, respondents with no computer experience or respondents who refused to take the assessment on the computer were given the paper-based assessment. According to the results of the background questionnaire, people who are directed to participate in this application in the computer environment must get enough points from two short tests, each of which takes 5 minutes to be completed, before they can take all the cognitive tests in three different domains in the computer environment. While the first of these two short tests measures basic ICT skills and the second one, namely cognitive pre-test, assess basic numeracy and literacy skills. Respondents who succeed in both tests participate in the application in a computer environment.

In the cognitive assessment tools, adults' skills in three domains such as literacy, numeracy and problem solving in technology rich environments were assessed on a computer-based environment. Specifically, literacy is defined as 'understanding, valuating, using and engaging with written texts to participate in society, to achieve one's goals, and to develop one's knowledge and potential' (OECD, 2019c, p.19). There are 58 literacy assessment items. The reliability of scores of the test regarding literacy ranges from .71 to .86 for the countries participated in Round 3 (OECD, 2019b).

Numeracy is defined as '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' (OECD, 2019c, p.24). There are 56 numeracy assessment items. The reliability of scores of the test regarding numeracy ranges from .69 to .90 for the countries participated in Round 3 (OECD, 2019b).

Problem solving in technology-rich environments is measured with 16 items and defined as 'using digital technology, communication tools and networks to acquire and evaluate information, communicate with others and perform practical tasks' (OECD, 2019c, p.28). The reliability of scores of the test regarding problem-solving ranges from .65 to .85 for the countries participated in Round 3 (OECD, 2019b).

Data Analysis

To show the applicability of survival analysis to correct response time data, domains literacy, numeracy and problem solving of the PIAAC data set of six countries were used and we applied the analysis in the Stata 15 program. The time until the correct response in these data sets were taken as survival time (t), and correct response was considered as failure ($\delta_i = 1$), while incorrect response and not responding the question within the survival time was taken as censored event ($\delta_i = 0$). Several covariates that are thought to affect the response time of the individuals who gave the correct answer are included in the model. In the problem-solving domain, the correct answer is taken as failure, while all options except the correct answer are considered as censored event. In addition, we removed the missing answers from the data sets in all domains of the countries. In the model, the first categories of the covariates such as gender, age, and educational attainment were taken as the reference category while the second categories of the covariates such as readiness to learn, the use of ICT skills at home, and the use of ICT skills at work were chosen as reference category. Choosing first categories of these covariates as reference categories would not make sense, as the first categories represented individuals who gave a zero answer to all items. KM survival estimates to examine the survival probabilities for the categories of covariates were obtained and the Tarone-Ware test was applied for all domains and countries to compare the survival probabilities of different categories of the covariates.

Graphical technique in which log-log survivor curves of the categories of each variable are plotted was used to determine whether the proportional hazard assumption was met for each country and domain. "A log-log survival curve is a transformation of an estimated survival curve, that derives from taking the natural log of the survival probability twice" (Dhoke et al., 2021, p.9). As an example, the corresponding graphs for the responses of the individuals in to the items assessing literacy domain were given in the Figure 1.
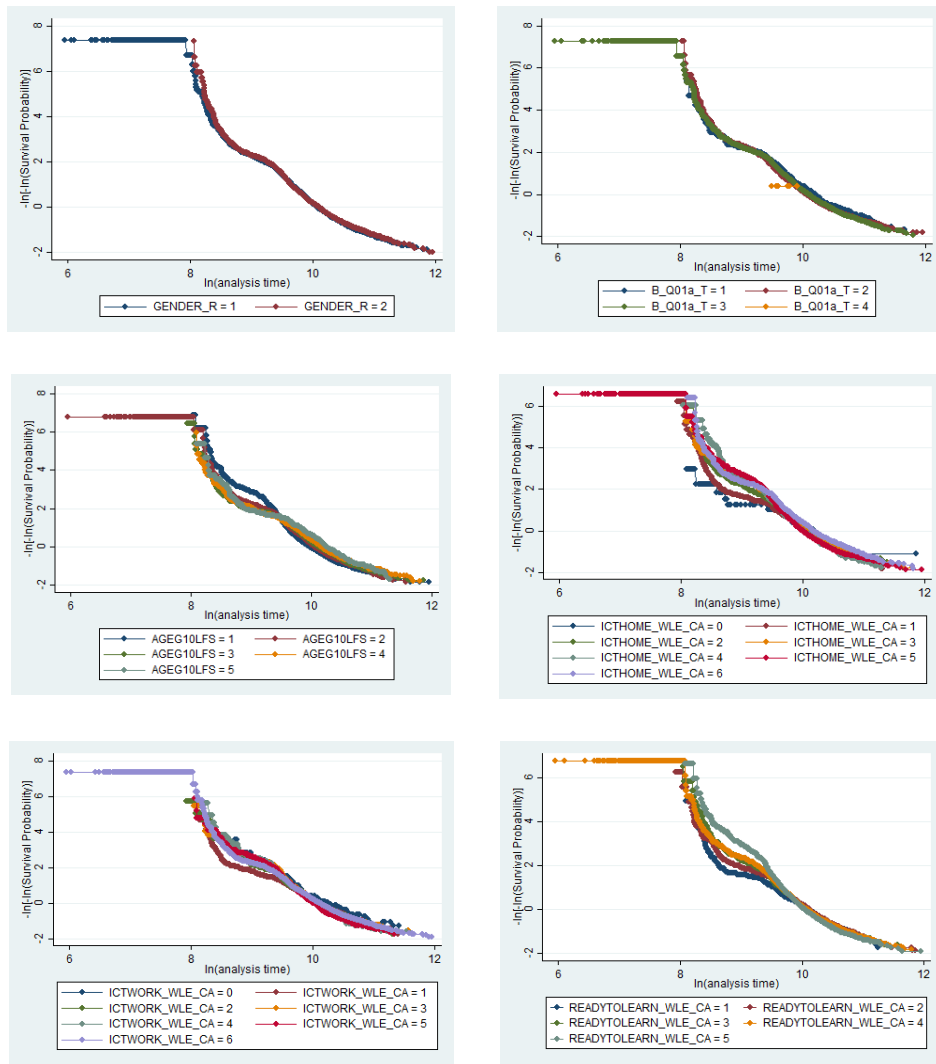
Figure 1. Log-log Survival Curves for Proportional Hazards Assumption of Literacy Domain for Peru

Accordingly, as it was depicted in Figure 1, proportional hazard assumption was violated because all categories of the covariates were intersecting to each other (Orbe et al., 2002). Similar results were seen for all other domains and countries. Since the proportional hazard assumption was violated for all covariates, the results of the proportional hazard models will not find to be satisfactory. Therefore, considering the applicability of AFT models to a greater number and variety of distributions than PH models (Saikia & Barman, 2017), AFT models for analysis the effect of covariates on survival time were used for a better interpretation of results. To check which AFT model suits the data best, Akaike Information Criterion (AIC) (Qiao et al., 2019) and Bayesian Information Criterion (BIC) (Weakliem, 1999) were used. AIC is represented in the equation below:

AIC = -2 (log likelihood) + 2(P +K)   (4)

where P is the number of parameters, and K is the number of coefficients (excluding constant) in the model. The lower value of AIC represents the best fit model (Dhoke et al., 2021). On the other hand, BIC is expressed as follows:

BIC = -2 (log likelihood) + (P +K)*ln(n)   (5)

where n is the number of observations. Similar to AIC, goodness of fit is represented by the lower BIC value (Dhoke et al., 2021).

The comparison results of the parametric AFT distribution models in terms of AIC and BIC values were presented in Table 3.

Table 3. Goodness of Fit of Different Models

Country	Domain		AIC	BIC	Country	Domain		AIC	BIC
Mexico	Literacy	**Weibull**	**424900.00**	**424963.33**	Hungary	Literacy	**Weibull**	**956731.20**	**956800.28**
		Log-normal	425128.40	425191.73			Log-normal	960948.00	961017.08
		Exponential	426988.20	427045.77			Exponential	961239.80	961302.60
		Log-logistic	426719.00	426782.33			Log-logistic	957969.80	958038.88
	Numeracy	Weibull	490198.80	490262.13		Numeracy	**Weibull**	**1104598.00**	**1104667.08**
		Log-normal	**487043.80**	**487107.13**			Log-normal	1116085.40	1116154.48
		Exponential	491921.60	491979.17			Exponential	1121731.60	1121794.40
		Log-logistic	489363.80	489427.13			Log-logistic	1105910.40	1105979.48
	Problem Solving	**Weibull**	**50233.40**	**50296.73**		Problem Solving	**Weibull**	**49436.60**	**49505.68**
		Log-normal	51731.60	51794.93			Log-normal	49844.20	49913.28
		Exponential	51235.20	51292.77			Exponential	50767.20	50830.00
		Log-logistic	51237.20	51300.53			Log-logistic	49622.60	49691.68
Peru	Literacy	Weibull	451276.60	451343.14	USA	Literacy	**Weibull**	**648103.00**	**648168.44**
		Log-normal	**447211.80**	**447278.34**			Log-normal	652588.80	652654.24
		Exponential	451274.60	451335.09			Exponential	667903.40	667962.89
		Log-logistic	456213.40	456279.94			Log-logistic	660453.00	660518.44
	Numeracy	**Weibull**	**560105.00**	**560171.54**		Numeracy	**Weibull**	**657238.60**	**657304.04**
		Log-normal	561222.80	561289.34			Log-normal	657447.40	657512.84
		Exponential	562872.00	562932.49			Exponential	677455.20	677514.69
		Log-logistic	566388.20	566454.74			Log-logistic	665041.20	665106.64
	Problem Solving	**Weibull**	**40182.60**	**40241.09**		Problem Solving	Weibull	88574.40	88639.84
		Log-normal	40982.60	41049.14			**Log-normal**	657238.60	657304.04
		Exponential	40937.60	40998.09			Exponential	677455.20	677514.69
		Log-logistic	40383.20	40449.74			Log-logistic	657447.40	657512.84
Kazakhstan	Literacy	Weibull	838979.20	839049.51	Ecuador	Literacy	Weibull	323738.00	323801.36
		Log-normal	**831797.80**	**831868.11**			**Log-normal**	**321570.60**	**321633.96**
		Exponential	844516.80	844580.72			Exponential	323744.40	323809.76
		Log-logistic	832892.40	832962.71			Log-logistic	322218.40	322281.76
	Numeracy	**Weibull**	**950470.00**	**950540.31**		Numeracy	**Weibull**	**376591.00**	**376654.36**
		Log-normal	952485.40	952555.71			Log-normal	377050.20	377113.56
		Exponential	962087.40	962151.32			Exponential	380946.60	381004.20
		Log-logistic	961235.40	961305.71			Log-logistic	379324.00	379387.36
	Problem Solving	**Weibull**	**98576.60**	**98646.91**		Problem Solving	**Weibull**	**24629.80**	**24693.16**
		Log-normal	99014.00	99084.31			Log-normal	25185.80	25249.16
		Exponential	99714.60	99778.52			Exponential	24949.80	25007.40
		Log-logistic	99716.40	99786.71			Log-logistic	24654.40	24717.76

Note. The best-fitted model is shown in boldface.

The results suggested that different AFT models suit the different data best.

Findings / Results

The results of the best fitted models obtained by examining the responses given by the adults from the countries participating in Round 3 of PIAAC to the items measuring three domains and the time they spend while giving correct answer to these items were given in Table 4.

Table 4. Results for All Domains

Covariate		Mexico			Peru			Kazakhstan			Hungary			USA			Ecuador		
		L	N	P.S	L	N	P.S	L	N	P.S	L	N	P.S	L	N	P.S	L	N	P.S
Cognitive Pre-test	-	-.15**	-.06**	-.06*	-.04**	-.17**	-21**	-.15**	-.21**	-.37**	-.15**	-.21**	-.19**	-.12**	-.10**	-.08**	-.11**	-.13**	-.10*
Attempt	-	.02**	.03**	.01**	-	.03**	-	.03**	.06**	.02**	.01**	.06**	.02**	.04**	-.05**	.01**	.04**	.04**	.01**
Gender (Female)	-	-	-	.07**	-	-	.06**	-	-	-	.20**	-.05**	.04**	-	-.13**	-	-	-.05*	.06**	-
Educational Attainment	2	-.15**	-.04**	-.25**	-	-.10**	-	-	-.09*	-	-	-	-	.06*	-	-	-	-.05*	-.19*
	3	-.15**	-.07**	-.30**	-	-.12**	-	-	-.14**	-	-	-.06*	-.26**	.07*	-	-	-	-	-
Age	2	.08**	-	-	-	-	-	-	-	-	.07**	-	.21*	.07**	-	-.39**	-	-.07*	.30**
	3	.11**	-.08**	.36**	-	-.06*	.18*	-	-	-	.15**	-	.50**	-	-	-	.09**	-.07*	.37**
	4	.21**	-.06*	.34**	-	-	-	-.06*	-	-	.21**	-	.72**	.15**	-	.29**	.15**	-	.76**
	5	.25**	-	.46**	-	-	.42**	-	-	-	.26**	-	.98**	.21**	.09**	.34**	.23**	-	.77**
Ready to learn	1	-	-	-	-.05*	-	.40**	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	.16**	-	-	-	-	.15**	.23**	.19*	.06**	.09**	-	-	-	-	-	-	-
	4	-	.14**	-	-	-	-	.16**	.23**	-	.05**	.10**	-	-	-	-	-	-	-
	5	-	.13**	-	.04*	-	-	.16**	.25**	-	.09**	.09**	-	-	.06**	-	-	-	-
	6	-	.10**	-	-	-	-	.16**	.26**	.28**	.12**	.08**	-	-	-	-	-	-	-
Use of ICT skills at home	1	-	-.11	-	-	-	-.71*	-	-.16**	-	-	-.21*	-	-	-.37**	-	-	-	-
	3	-	-	-	-	-	-	-	-	.23**	-.08**	-.11**	-.13*	-.06*	-	-.15*	-.07*	-	-
	4	-.07*	-	-.19*	-	-	-.26**	-	-.08*	-	-.07**	-.07**	-.12*	-.10**	-	-.18**	-	-	-
	5	-.10**	-	-.39**	-	-	-.29**	-	-.09*	-	-.07**	-.09**	-.23**	-.13**	-	-.19**	-	-	-
	6	-.14**	-	-.16*	-	-	-.26**	-	-.20**	-	-.12**	-.06**	-.29**	-.12**	-	-.26**	-.10**	-	-
Use of ICT skills at work	1	-	-	-	-	-	-	-	-.17*	-	-.07*	-	-	-	-	-	-	-	-
	3	-	-	-	-	-.08*	-	-	-	-	-.04	-	-	-.08**	-.07**	-.12*	-	-	-.31*
	4	-	-	-	-	-	-	-	-	-	-.15**	-.11**	-.20**	-.13**	-.11**	-.22**	-	-	-.35*
	5	-.15**	-	-.21*	-	-.13**	-	-	-	-	-.16**	-.06**	-.25**	-.21**	-.12**	-.32**	-	-.12**	-.46**
	6	-.16**	-	-.26**	-	-.11**	-	-	-	.56**	-.14**	-.09**	-.32**	-.17**	-.09**	-.32**	-	-.15**	-.44**

Note.$*p<.05$. $**p<.01$. L= Literacy. N= Numeracy. P.S= Problem Solving.

As stated in Table 4, a significant negative relationship was found for all countries included in the scope of the study between the scores obtained by adults from the cognitive pre-test and the time they spent to give correct response to the items that measure literacy skills (p<.01). Accordingly, as the scores of cognitive pre-test of adults in USA, Mexico, Peru, Kazakhstan, Hungary and Ecuador increased, response time for giving the correct answer decreased. Specifically, each unit increase in scores obtained from cognitive pre-test of the adults in Mexico and Hungary caused a decrease the response time to give correct answer by 14% (exp(-.15) =.86). In other words, it was found that individuals who got high scores from the test assessing the basic computer skills gave the correct answer in a shorter time. The similar patterns were also observed for the domains of numeracy and problem solving.

Another finding of this study is that as the number of attempts increases, the response time of the adults who gave the correct answer for the items in all domains increased (p<.01). Specifically, it was found that each unit increase in the number of attempts performed by the individuals in Mexico, Kazakhstan, Hungary, USA and Ecuador increases the response time passed until the individuals give correct answer for the items assessing literacy skills by 2% (exp (.02) = 1.02), 1% (exp (.01) = 1.01), and 3% (exp (.03) = 1.03), respectively. Except for USA, a significant positive correlation was also found between the number of attempts exhibited by adults in other countries and response time for the ones who gave the correct answer for the items measuring skills regarding numeracy and problem solving (p<.01). On the other hand, each unit increase in the number of attempts performed by the individuals in USA leaded to decrease in the response time by 5% (p<.01).

The statistically significant relationship of the gender with the time taken to answer correctly showed that response time for giving correct answer to the items measuring the literacy items of women in Hungary and Ecuador is .95 times that of men (exp (-.05) = .95), while it is .88 times that of men (exp(-.13)=0.88) in USA. On the other hand, the fact that women give correct answers in a shorter time than men do not apply to the numeracy and problem solving domains. Specifically, as compared to men, women in Mexico, Peru, Hungary, and Ecuador spent 7% (exp(.07) = 1.07-1 = .07), 6%, 4%, and 6% more time to answer items that measure numeracy skills, respectively. For the domain of problem solving, women's response time for giving correct answers is 22% more than men's response time (1.22-1.00 = .22).

As it was presented in Table 4, the relationships between the time spent by adults in Mexico and USA to answer literacy items and educational attainment levels were found to be statistically significant (p<.01). For the three domains, a general pattern was found that there was a negative relationship between response time for giving correct answer and educational attainment of adults. To give an example for the findings of the literacy domain, in Mexico, individuals whose education level is above high school or high school have %14 decrease in the response time to give correct answer compared to individuals whose education level is less than high school. To put it indifferent words, response time of individuals whose education level is above high school or high school is .86 times that of individuals whose education level is lower than high school (exp(-.15) = .86). In America, the opposite result was found. In other words, the response time of individuals with an education level of high school or above was found to be 1.06 times that of those with an education level below high school (exp(.06) = 1.06).

As it can be seen in Table 4, for the literacy and problem solving domains, the time relatively older individuals aged above 24 in Mexico, Peru, Hungary, Ecuador spent until they gave answer the item was longer than that of younger individuals who aged under the age of 24. For example, in Mexico, the response times of adults aged 55 and over to correctly answer the items measuring literacy skills are 1.28 times that of individuals aged 24 and under (exp (.25) = 1.28). In other words, individuals aged 55 and over spend 28% more time (1.28- 1.00 = .28) to answer an item correctly than individuals aged 24 and under. On the other hand, the response times of adults aged 45-55 to correctly answer items assessing literacy skills in Kazakhstan decreased that of individuals aged 24 and under (p<.05). When looking at the general picture of the relationship between the age of individuals in the USA and the time they spend to answer correctly to items assessing problem-solving skills, it can be stated that there is a similar situation in this country. However, there is an exceptional case for individuals between the ages of 25-34. To be more precise, individuals in this age range spend less time answering items correctly than individuals aged 24 and under. For the numeracy domain, the findings were differed from the findings obtained from literacy and problem-solving domains. That is, older adults in Mexico and Peru were found to respond to items assessing skills regarding numeracy faster than younger adults.

As can be seen from Table 4, for all domains, in most of the countries included in this study, response time for answering the items of the individuals whose readiness to learn strategies use rate in different situations is above 20% was longer than that of the individuals whose strategy use rate is less than 20%. For example, in Kazakhstan, the time spent by adults who use readiness to learn skills at 80% rate, to answer items that measure numeracy skills correctly is 1.29 (exp (.26) = 1.29) times that of individuals who use these skills at 20% frequency. In other words, response time of adults who used these skills intensively and gave the correct answer was 29% more than that of individuals who use them rarely.

As stated in Table 4, for all domains and most of the countries included in the study, response time for answering the items of the individuals who used ICT skills at home or work at least 20% rate was shorter than that of the individuals whose rate of ICT use was less than 20%. For example, in Hungary, the response times of adults who use the computer in home environments at 80% rate and give the correct answer are .89 times that of individuals whose rate of ICT skill

use in home environment less than 20% (exp (-.12) = .89). A similar situation is valid for adults in Mexico and Ecuador (*p*<.01). Additionally, when the relationship between the frequency of using ICT skills of these individuals in the workplace and their response times for giving correct answer was examined the same pattern was found. Specifically, it was found that the more often individuals in Mexico, Hungary and USA use their ICT skills in their workplace, the more likely they gave answer quickly to the item compared to individuals whose rates of ICT skills use is 20%. In general, although a shorter response time was found to be associated with the frequent use of ICT skill, Kazakhstan displayed as an exceptional case for the domains of numeracy and problem solving. A significant positive relationship was found between the frequency of using ICT skills of adults both at home and in their work life and the time spent to give correct answer. In other words, it was observed that response times of high ICT users were longer than low ICT users.

Discussion

Once analyses were performed with AFT models, it was concluded that individuals who score higher in cognitive pre-test tend to give correct answers more quickly. The reason for this situation may be that those who score higher in the cognitive pre-test also get enough scores from a short pre-test that measures their ability to use basic ICT tools and skills given before this test (Kirsch et al., 2013). To be more precise, considering that only individuals participating in the computer-based application of the PIAAC are included in the research, it is expectable that basic ICT skills correlate with the performance in such computer-based tests. This situation can be explained by the possibility of increasing ICT familiarity levels depending on the development of basic ICT skills (Odo, 2012; Pomplun et al., 2002). In addition, these individuals, who have basic literacy and numeracy skills, may be more likely to know the answers to the questions in the items presented to them in all three domains, which may have led them to the correct answer in a shorter time.

For most countries participating in the Round 3 of the PIAAC, a positive correlation was observed in all domains between the number of attempts displayed by adults and the response time of the adults who gave the correct answer to items. In other words, as the number of attempts made to answer an item increases, it has been observed that the correct response times for the items increase. The high number of attempts may be an indication that individuals exhibit solution behavior and, accordingly, they are more engaged to obtain the correct answer (Sahin & Colvin, 2020). Therefore, they may have spent more time answering the items (Schnipke & Scrams, 1997). On the other hand, it has been observed that as the number of attempts displayed to answer the items regarding the numeracy domain in the United States increases, the time to answer the item correctly decreases. This differentiation in the USA regarding the relationship between the attempt number and the time spent to answer the item correctly may have resulted from the USA's familiarity with computer-based applications. To be more precise, the USA has a large number of computer based assessment practices (i.e, Measures of Academic Progress, Test of English for International Communication, The Praxis Series: Teacher Licensure and Certification, Preprofessional Skills Test, etc.) compared to other countries participating in Round 3. Therefore, adults in this country are more likely to participate in such practices at some time in their lives and have experience in participating such computer-based assessments (Scheuermann & Bjornsson, 2009).

In this study, it was concluded that the time spent by women and men in most countries to correctly answer the literacy, numeracy and problem solving items differ statistically from each other. To be more precise, the time spent by women to correctly answer numeracy and problem solving items that require numerical skills is longer than men, while it is shorter for items requiring verbal skills. Similar findings were found in another study conducted by Goldhammer et al., (2017), which examined the differentiation between test disengagement of women and men. Researchers have found that women spend more time responding to items that require them to use problem solving skills than men. The reason may be because women are more successful in tests that require verbal skills and therefore reach the correct answer faster (Balart & Oosterveen, 2019).

In this study, individuals with a higher educational attainment level reach the correct answer in a shorter time compared to individuals with an educational attainment level less than high school. This situation can be explained by that individuals with a low level of education cannot receive formal education enough to develop literacy, numeracy and problem solving skills, and therefore have a lower competency level (Massing & Schneider, 2017), and work in relatively simple jobs (Kureková et al., 2013), have the lower computer experiences (OECD, 2015). However, in the literacy domain, the opposite of this situation in the USA has been observed. That is, as the education level increases, the time to respond correctly to the items increases. It is thought that this may be due to a statistical error, which means that the significance level for most countries is 99% for the educational level covariate, while for the USA it is 95%.

When the relationships between the age of the individuals and the time to answer the items were analyzed, the general pattern observed in most countries that the older the individuals, the slower they answered the literacy and problem solving items. This differentiation in the response times of individuals and accordingly in their response speed is more pronounced, especially in adults aged 55 and over. This situation is thought to be due to the cognitive aging of individuals, slowing down in information processing and reading comprehension (Ratcliff et al., 2004) and their low level of familiarity with computers (Harvey et al., 2020). However, in the numeracy domain, it has been observed that in most countries, individuals between the ages of 35 and 54 respond to the items in a shorter time than individuals

under 24. This may be due to the experiences of individuals between the ages of 35-54 and the knowledge they have gained compared to individuals under the age of 24.

In most of the countries included in the study, it was concluded that the individuals spent longer time solving the item by using readiness to learn strategies instead of giving correct answers quickly. This situation may be related to the perseverance and openness to problem solving levels of the individuals. More clearly, individuals with high perseverance or openness to problem solving levels are willing to answer the items given to them, seek other ways to solve the item even if they do not know the answer, and tend to be more engaged (OECD, 2013c). This situation may have caused them to spend more time. A similar finding was found in the research conducted by Ilgun Dibek (2020). The researcher stated that as individuals' readiness to learn levels increase, their engagement levels also increase.

In nearly all of the countries included in the study, it was concluded that individuals who use ICT skills frequently give correct answers to the items more quickly. Since the items measuring literacy, numeracy and problem solving skills are given to individuals in the computer environment, it is important that individuals should have these skills in order to give the correct answer, as well as using ICT skills effectively. Therefore, due to their frequent use of ICT skills, their familiarity with ICT-related technological equipment and their increased experience may have enabled individuals to answer the items more quickly (Pomplun et al., 2002).

Conclusion

The main purpose of this article is to demonstrate the applicability of survival models, which are frequently used in fields such as medicine and business, to educational data by using data set of countries participated in Round 3 of PIAAC on literacy, numeracy and problem solving domains. For this, various covariates have been included in survival models. As a result of the analysis, it was concluded that the relationship between the covariates and the time spend on answering the items correctly vary from domain to domain and from country to country.

Recommendations

Considering that response time is an important evidence in determining the individuals' test dis/engagements which cause false estimation of test ability levels of individuals by mixing measurement invariance with tests, inferences and practices made at the educational policy level in international practices will create serious negative consequences when response time of individuals are ignored. Examining the factors associated with item responses is valuable in predicting and understanding them correctly. For example, if educational policy makers, who make new regulations based on the difference between groups in general and between the ability levels of women and men in particular, do not take into account the differences in the response times of individuals, the validity of their practices will be low. At the same time, investigating item response times is valuable both in determining the effort of individuals regarding the test and how much effort should be made on which item. For this reason, it is recommended for practitioners to make a routine process of examining test taking efforts, which provide important information about the performance of individuals in testing programs that use linear online tests or computer adaptive tests.

Limitations

There are some limitations regarding the research. One of the limitations of this study is that the results of the study cannot be generalized to all countries. To be more precise, although all the countries in Round 3, of the PIAAC, were included in the study, countries participating in different rounds of the PIAAC were not included in this study. Therefore, other studies can be conducted to examine the applicability of survival models by addressing more countries. In addition, by examining the parameter estimation powers of survival models and other latent traits models used when modeling PIAAC data, it can be determined which model gives more unbiased parameter estimation results. In addition, since the main purpose of this study is to demonstrate the applicability of survival models to educational data and especially to response time data, the number of covariates that may be associated with individuals' correct response times was limited to six. Therefore, further studies can be conducted to examine the relationship between more and different variables with the correct response time. Another limitation is that since the survival models analyzed in this study are basically based on regression, it can be interpreted whether covariates predict the correct response time, which is the dependent variable, while a causality relationship cannot be established between these variables.

References

Anghel, B., & Balart, P. (2017). Non-cognitive skills and individual earnings: New evidence from PIAAC. *Journal of the Spanish Economic Association, 8*(1), 417–473. https://doi.org/10.1007/s13209-017-0165-x

Balart, P., & Oosterveen, M. (2019). Females show more sustained performance during test-taking than males. *Nature Communications, 10*(1), 3798. https://doi.org/10.1038/s41467-019-11691-y

Bekalo, D. B. (2019). Modeling determinants of time-to-circumcision of girls: a comparison of various parametric shared frailty models. *Health Services and Outcomes Research Methodology, 19*, 145–174. https://doi.org/10.1007/s10742-019-00199-z

Chernick, M. R., & Friis, R. H. (2003). *Introductory biostatistics for the health sciences: Modern application including bootstrap*. John Wiley and Sons.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B, 34*(2), 187-220. https://www.jstor.org/stable/2985181

De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Applied Mathematics and Statistics, 10*, 102. https://doi.org/10.3389/fpsyg.2019.00102

Dhoke, A., Kumar, A., & Ghosh, I. (2021). Hazard-based duration approach to pedestrian crossing behavior at signalized intersections. *Transportation Research Record*. Advance online publication. https://doi.org/10.1177/03611981211003102

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153.,

Finn, B. (2015). *Measuring motivation in low-stakes assessments* (Research Report No. RR-15-19). Princeton. https://doi.org/10.1002/ets2.12067

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. SAGE.

Goldhammer, F., Martens, T., & Lüdtke, O. (2017). Conditioning factors of test-taking engagement in PIAAC: An exploratory IRT modelling approach considering person and item characteristics. *Large-scale Assessments in Education, 5*(18), 1-25. https://doi.org/ 10.1186/s40536-017-0051-9

Goldhammer, F., Naumann, J., Stelter, A., Toth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*(3), 608–626. https://doi.org/10.1037/a0034716

Hart, T., Chaparro, B., & Halcomb, C. (2008). Evaluating websites for older adults: Adherence to senior-friendly guidelines and end-user performance. *Behaviour & Information Technology, 27*(3), 191–199, https://doi.org/10.1080/01449290600802031

Harvey, P. D., Tibiriçá, L., Kallestrup, P., & Czaja, S. J. (2020). A Computerized functional skills assessment and training program targeting technology based everyday functional skills. *Journal of Visualized Experiments*, 156, 1-18, https://doi.org/10.3791/60330

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*(150), 1-19. https://doi.org/10.3389/fnins.2014.00150

Ilgun Dibek, M. (2020). Silent predictors of test disengagement in PIAAC 2012. *Journal of Measurement and Evaluation in Education and Psychology, 11*(4), 430-450. https://doi.org/10.21031/epod.796626

Kirsch, I. S., Lennon, M. L., von Davier, M., Gonzalez, E. J., & Yamamoto, K. (2013). *On the growing importance of international large-scale assessments, in the role of international large-scale assessments: Perspectives from technology, economy, and educational research*. Springer. https://doi.org/10.1007/978-94-007-4629- 9_1

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text*. Springer.

Kureková, L., Haita, C., & Beblavý, M. (2013). Conceptualizing low-skillness: A new approach. *Sociology, 45*(3), 247-266.

Massing, N., & Schneider, S. L. (2017). Degrees of competency: The relationship between educational qualifications and adult skills across countries. *Large-scale Assessment in Education, 5*(1), 1-6. https://doi.org/10.1186/s40536-017-0041-y

Nardi, M., & Schemper, M. (2003). Comparing Cox and parametric models in clinical studies. *Statistics in Medicine, 22*, 3597-3610. https://doi.org/10.1002/sim.1592

Novikov, N. A., Nurislamova, Y. M., Zhozhikashvili, N. A., Kalenkovich, E. E., Lapina, A. A., & Chernishev, B. V. (2017). Slow and fast responses: Two mechanisms of trial outcome processing revealed by EEG oscillations. *Frontiers in Human Neuroscience. 11*, 218. https://doi.org/10.3389/fnhum.2017.00218

Odo, D. M. (2012). Computer familiarity and test performance on a computer-based cloze ESL reading assessment. *Teaching English with Technology, 12*(3), 18-35. https://eric.ed.gov/?id=EJ1145240

Orbe, J., Ferreira, E., & Nunez-Anton, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine, 21*, 3493-3510. https://doi.org/10.1002/sim.1251

Organization for Economic Co-operation and Development. (2013a). *OECD skills outlook 2013: First results from the survey of adult skills*. OECD Publishing. https://doi.org/10.1787/9789264204256-en

Organization for Economic Co-operation and Development. (2013b). *Skilled for life? Key findings from the survey of adult skills,* OECD Publishing. https://bit.ly/2YxcaIN

Organization for Economic Co-operation and Development. (2013c). *PISA 2012 Results: Ready to learn: Students' engagement, drive and self-beliefs* (Volume III), PISA. OECD Publishing. https://bit.ly/3Fwyljo

Organization for Economic Co-operation and Development. (2015). *Adults, computers and problem solving: What's the problem?* OECD Publishing. https://doi.org/10.1787/9789264236844-en

Organization for Economic Co-operation and Development. (2019a). *Beyond Proficiency: Using log files to understand respondent behaviour in the survey of adult skills, OECD Skills Studies*, OECD Publishing. https://doi.org/10.1787/0b1414ed-en

Organization for Economic Co-operation and Development. (2019b). *Technical report of the survey of adult skills (PIAAC)*, OECD Publishing. https://bit.ly/2YzyfGU

Organization for Economic Co-operation and Development. (2019c). *The survey of adult skills: Reader's companion* (3rd ed.). OECD Publishing. https://doi.org/10.1787/f70238c7-en

Organization for Economic Co-operation and Development. (2020). *Education at a glance 2020: OECD indicators*, OECD Publishing. https://doi.org/10.1787/69096873-en

O'Neil, H. F., Sugrue, B., & Baker, E. L. (1995). Effects of motivational interventions on the national assessment of educational progress mathematics performance. *Educational Assessment*, *3*(2), 135–157. https://doi.org/10.1207/s15326977ea0302_2

Pomplun, M., Frey, S., & Becker, D. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and Psychological Measurement, 62*(2), 337-354. https://doi.org/10.1177/0013164402062002009

Qiao, Y., Labi, S., & Fricker, J. D. (2019). Hazard-based duration models for predicting actual duration of highway projects using nonparametric and parametric survival analysis. *Journal of Management in Engineering, 35*(6), 1-16. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000700

Qui, J. (2009). *Comparison of proportional hazards and accelerated failure time models* [Unpublished master's thesis]. University of Saskatchewan.

Quinn, D. M., & Cooc, N. (2015). Science achievement gaps by gender and race/ethnicity in elementary and middle school trends and predictors. *Educational Researcher, 44*(6), 336–346. https://doi.org/10.3102/0013189X15598539

Ranger, J., Kuhn, J. T., & Ortner, T. M. (2020). Modeling responses and response times in tests with the hierarchical model and the three-parameter lognormal distribution. *Educational and Psychological Measurement, 80*(6), 1059–1089. https://doi.org/10.1177/0013164420908916

Ranger, J., & Ortner, T. (2012). A latent trait model for response times on tests employing the proportional hazards model. *British Journal of Mathematical and Statistical Psychology, 65*(2), 334–349. https://doi.org/10.1111/j.2044-8317.2011.02032.x

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19*(2), 278–289. https://doi.org/10.1037/0882-7974.19.2.278

Reder, S. (1998). In MC. Smith (Ed.). *Literacy for the 21st century: Research, policy, practice and the National Adult Literacy Survey*. Greenwood Press.

Rouder, J., Province, J., Morey, R., Gomez, P., & Heathkote, A. (2015). The lognormal race: a cognitive- process model of choice and latency with desirable psychometric properties. *Psychometrika, 80*(2)*,* 491–513. https://doi.org/10.1007/s11336-013-9396-3

Sahin, F., & Colvin, K. F. (2020). Enhancing response time thresholds with response behaviors for detecting disengaged examinees. *Large-scale Assessment in Education, 8*(5). https://doi.org/10.1186/s40536-020-00082-1

Saikia, R., & Barman, M. P. (2017). A review on accelerated failure time models. *International Journal of Statistics and Systems, 12*(2), 311-322. https://bit.ly/3Btzw0o

Sawaki, Y. (2001). Comparability of conventional and computerized tests of reading in a second language. *Language Learning & Technology*, *5*(2), 38-59. http://dx.doi.org/10125/25127

Scheuermann, F., & Bjornsson, J. (Eds.). (2009). *The transition to computer-based assessment.* Office for Official Publications of the European Communities.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement, 34*(3), 213-232. https://doi.org/10.1111/j.1745-3984.1997.tb00516.x

Singh, R., & Mukhopadhyay, K. (2011). Survival analysis in clinical trials: Basics and must know areas. *Perspectives in Clinical Research*, *2*(4), 145–148. https://doi.org/10.4103/2229-3485.86872

Smith, M. C., Rose, A. D., Ross-Gordon, J., & Smith, T. J. (2015). *Adults' readiness to learn as a predictor of literacy skills*. https://bit.ly/3lnVlc9

Sternberg, R. J. (1985). *Beyond IQ. A triarchic theory of human intelligence*. University Press.

Swindell, W. R. (2009). Accelerated failure time models provide a useful statistical framework for aging research. *Experimental gerontology, 44*(3), 190-200. https://doi.org/10.1016/j.exger.2008.10.005

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, *72*(3), 287–308. https://doi.org/10.1007/s11336-006-1478-z

van der Maas, H. L. J., Molenaar, D., Maris, G., Kievit, R., & Borsboom, D. (2011). Cognitive psychology meets psychometric theory: On the relation between process models for decision making and latent variable models for individual differences. *Psychological Review, 118*(2). 339–356. https://doi.org/10.1037/a0022749

van Zandt, T. (2002). Analysis of response time distributions. In H. Pashler & J. Wixted (Eds.), *Stevens' handbook of experimental psychology: Methodology in experimental psychology* (pp. 461–516). John Wiley & Sons Inc.

Wang, C., Fan, Z., Chang, H. H., & Douglas, J. A. (2013). The linear transformation model with frailties for the analysis of item response times. *British Journal of Mathematical and Statistical Psychology, 66*(1), 144–168. https://doi.org/10.1111/j.2044-8317.2012. 02045.x

Wang, S., & Chen, Y. (2020). Using response times and response accuracy to measure fluency within cognitive diagnosis models. *Psychometrika, 85*(3), 600–629. https://doi.org/10.1007/s11336-020-09717-2

Weakliem, D. L. (1999). A critique of the Bayesian information criterion for model selection. *Sociological Methods & Research, 27*(3), 359–397.

Wei, L. J. (1992). The accelerated failure time model: A useful alternative to the Cox regression model in survival analysis, *Statistics in Medicine*, *11*(14-15), 1871-1879. https://doi.org/10.1002/sim.4780111409

Wenger, M., & Gibson, B. (2004). Using hazard functions to assess changes in processing capacity in an attentional cuing paradigm. *Journal of Experimental Psychology*, *30*(4), 708–719. https://doi.org/10.1037/0096-1523.30.4.708

Wheatley-Price, P., Hutton, B., & Clemons, M. (2012). The Mayan Doomsday's effect on survival outcomes in clinical trials. *Canadian Medical Association Journal, 184*(18), 2021–2022. https://doi.org/10.1503/cmaj.121616

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wise, S. L., & DeMars, C. E. (2010). Examinee noneffort and the validity of program assessment results. *Educational Assessment*, *15*(1), 27–41. https://doi.org/10.1080/10627191003673216

Wise, S. L., & Kong, X. J. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163–183. https://doi.org/10.1207/s15324818ame1802_2