Chapter 3: Methods

3.1 Data

The international dataset used in this analysis is secondary data from the student level OECD's PISA 2018 cycle provided for public at OECD official website. In addition to data from the main cognitive test, data from both the ICT Familiarity Questionnaire (SC) and the Students Background Questionnaire (SC) were also included in the analysis to obtain contextual knowledge about students. In its latest cycle, 79 countries participated in PISA assessment, as an optional questionnaire, only 52 countries have administered the ICT questionnaire to students. (SC)

Using the 4.0.3 version of *R* software and *intsvy* package, the original data file from OECD was treated to produce a final dataset including the three Scandinavian countries with total sample size of $N = 18810$ observations —by observations we refer to the 15-year-old students who participated in the assessment— distributed as 7657 Denmark, 5649 Finland, and 5504 students from Sweden. Participants were chosen randomly within their own cluster i.e. school. With 348 participating school from Denmark, 214 from Finland and 223 from Sweden, the total number of schools was 785. It is worth to mention that the items used for the cognitive tests, including reading, are confidential except for small numbers that were released to demonstrate the scaling process (SC).

3.2 Instruments

Table (1) exhibits an overview of the instruments (variables) used in this study; these items are further explained below.

3.2.1 ICT Usage: The Predictors (X)

ICT use-related independent variables are drawn from the ICT Familiarity Questionnaire (SC), for this study, three indices were used to capture the entire scope/realm of students' usage of ICT devices. These continuous indices are derived—by means of IRT scaling technique—from items that investigates the students' place, purpose, and frequency of this usage, i.e. where, what for and how often they use the ICT.

These items, rather activities, build the indexes in question, namely, (a) *USESCH* the students general use of ICT at school e.g. browse the Internet for schoolwork and doing homework on a school computer (10 items), (b) *ENTUSE* the use of ICT outside the school for leisure e.g. playing one-player games and chatting online, (12 items), and (c) HOMESCH use of ICT outside the school for school-work e.g. browsing the internet to follow up lessons and using social networks for communication with teachers (11 items).

All the items are self-reported and correspond to a 5-point Likert scale about the frequency of each activity (item), ranging from 1 = "never or hardly ever" to 5= "everyday", (see app.X for the detailed items). OECD standardized these indices to a mean of 0 and *SD* of 1, therefore higher values of theses indices indicate more frequency of the activity (SC).

3.2.2 Reading Attitudes: The Mediators (M)

Following the Theory of Planed Behaviour (TPB), three reading related-attitudes indices from PISA student background questionnaire (APPX) were implemented to correspond to the three core components of the TPB as shown in Figure 1 above, namely, attitude, subjective norm and perceived behavioral control (SC).

The three indices, or indicators, are derived variables from numbers of items. The READJOY variable measures the enjoyment of reading—corresponds to the attitude component— by using 5 self-reporting items e.g. “I read only if I have to” and “reading is one of my favorite hobbies” with four Likert-scale response categories ranging from “strongly disagree” to “strongly agree”. The items of this variable were mixed (negatively and positively) worded, therefore the negatively worded items were reverse-scored to implement IRT scaling so that higher values indicate greater enjoyment of reading (SC).

The second variable, READCOMP—corresponds to the Subjective Norm component—is a 3-items derived indictor that measures the student’s self-concept of reading competence, in other words their perception of competence in reading, through 4 Likert scale categories, strongly disagree to strongly agree, all the 3 items are positively worded such as “I am a good reader” therefore higher values indicate greater self-concept of competence.

Finally to represent the third component of the TPB, the Perceived Behavioral Control, the variable measuring student’s self-concept of reading difficulty was used. The variable is constructed from 3 items that are negatively worded e.g. “I have to read a text several times before completely understanding it”. Since the other two indicators were positively worded, and in order to avoid mixed signals, this variable was reverse-coded so that higher values indicate greater self-conception of ease, not difficulty, while reading. Finally, as Table 2 shows, the scale reliabilities for all the independent variables range between acceptable and excellent, as reported by OECD (TECH REPORT).

3.2.3 Reading Performance: The Outcome (Y)

Reading, the dependent variable in this study, was the main domain of PISA 2018 assessment, therefore more test items focused on reading abilities (245 items) than the other two domains i.e. science (115 items) and mathematics (83 items), this new inclusion of larger number of reading items improved measurement.

To further increase measurement validity and accuracy, computer-based multistage adaptive method was used for reading test, in that students receive the next items based on their performance in the ones before (SC). The 2018 test lasted for two hours and was mainly computer-based. The reading items varied in difficulty level and type including multiple-choice and short-constructed. According to OECD, its reading framework involved several cognitive processes (Table 3) and the distribution of tasks, or marks, in the reading test corresponded to these cognitive processes (Table 4).

Since PISA uses rotated-booklet design, a number of 15 clusters (units) of items was developed. Students were not administered the complete set of reading items, instead they responded to 7 units—between 33 and 40 items— depending on which testlet was taken at each stage (SC). This "missing by design" is the reason why PISA is not able to provide one single measure for cognitive domains such as reading, rather, PISA estimates students cognitive abilities using 10 plausible values (PVs) calculated by multiple imputation from both reading test and background questionnaire (SC, von Davier 2014, SC), these PVs represent the range of abilities of a student if they had completed the whole test, SC). As per prior recommendations from (OECD, 2009; Rutkowski et al., 2010;) all the ten PVs of reading (PV1READ – PV10READ) were included, one data set for each PV, then combining all parameters following Rubins combination (SC) to better measure parameters of the mediation model. Finally it is important to mention that, as the case for the other cognitive scales, reading PVs were set with a mean of 500 and a standard deviation of 100.

3.2.4 Student Background: The Control Variables

In addition to the ICT, reading attitudes, and reading performance, three student demographic variables were included in the analysis to serve as controlling variables and to draw groups comparison. The economic, social and cultural status (ESCS), an index derived from a) parents' education level, b) parents' occupation and c) home possessions, to indicate social and economic status of students (SC). When it comes to SEM, scholars argue whether to measure ESCS in a reflective or formative measurement model (SC), however this study used the OECD scale, i.e. formative measurement.

The second demographic variable is related to immigration status, whether the student is a native, second generation (IMMI2) or a first generation(IMMI1) in the test country. This information was derived from two items (parent's country of birth) and (students age of arrival in test country). Finally, student's gender was also included and recoded a dummy variable where female = 0 and male = 1.

3.3 Data Analysis

This study used full information maximum likelihood (FILM) method under the assumption that they occurred randomly (Enders, 2010), a method well suited to handle data missing at random (SC). All the missing values were recoded to (-99) with the purpose that Mplus software treats them with FIML method (SC). An SPSS(SC) diagnose for variance inflation factor (VIF) showed that no variable poses multicollinearity threat, as all VIF values were <5 appX (SC).

Since PISA uses stratified *random* sampling in all the participating countries (SC), in which schools were sampled at the first stage, and students within schools were sampled in the second stage, such design will lead to different probabilities for students to be chosen as participants in the assessment(SC). To ensure data is nationally and internationally representative, OECD employs weighting where the final student weight include both “school weight (the inverse of the school’s probability of selection) and the within-school student weight (the inverse of the student s probability of selection)” (SC).

Using these replicate weights in analyses would reduce the sampling bias (SC) as this study does by the virtue of the *WEIGHT* command in Mplus (SC). Finally, by default Mplus implements a robust maximum likelihood estimator (MLR), a widely used estimator to deal with the type of data this study used (SC). The command CLUSTER in Mplus was used to account for the nesting i.e. students nested in schools, further the COMPLEX type of analysis was chosen to obtain corrected standard errors and chi square test of model fit while considering stratification and non-independence of observations (SC)(SC).

3.3.1 Structural Equation Modelling

"Structural equation modeling can be defined as a class of methodologies that seeks to represent hypotheses about the means, variances, and covariances of observed data in terms of a smaller number of 'structural' parameters defined by a hypothesized underlying conceptual or theoretical model."

(Kaplan, 2001)

The intended model in this study is structural equation modeling (SEM), a modeling that is generated from combining two components, namely, measurement (factor analysis) and structural (path analysis) (SC).

3.3.2 Factor Analysis

Inasmuch as PISA applies IRT (item response theory) approach to measure latent constructs (SC) such those used in this study, it is appropriate to conducted a secondary a) confirmatory factor analysis (CFA) to validate the construct(s) of the independent variables (SC) and b) principal component analysis (PCA) and exploratory factor analysis (EFA) to identify the distinct latent construct(s) underlying the ICT use at school (USESCH) observed variables as several literature indicates that this usage is—as in outside school context— can be for leisure or schoolwork (SC).

Taking into consideration that the items are of categorical/ordinal nature, it is recommended to choose robust weighted least squares mean and variance (WLSMV) estimator when conducting measurement models (SC). Lavaan package (SC) of R (SC) provides this estimator by choosing estimator = WLSMV. Finally, this study applies oblique rotation in EFA, an appropriate rotation for correlated factors (SC), which Psych package provides (SC).

3.3.3 Path Analysis

After assessing the measurement part, it is rational to analyze the structural part. To specify structural model it is constructive to describe the model using path diagrams first introduced by (SC). Path coefficients, manifested in Figure 2, specify and assess the relationships between a) predictors (X) and mediators (M), b) mediators and outcome variable (Y), c) predictors (X) and outcome variable (Y). In the diagram, path coefficients β_{44}, β_{54}, and β_{64} have direct effect on the outcome READ, while all the other path coefficients have indirect effect on the outcome. The indirect effect is the product of both coefficients of the path i.e. (a*b) as illustrated in Figure 3 below (SC).

By using MODEL INDIRECT command, Mplus will, by default, implement multivariate delta method (SC)suggested by (Sobel, 1982) which produce more accurate results for large samples (SC). In correspondence with the suggested mediation effect in chapter 2, this analysis adopted mediation model (SC) following the equations in Figure 4 for the path analysis in the model, to elaborate the equations Figure 5 shows the matrix of the model.

3.4 Model Modification and Evaluation

First, considering the clustering nature of the PISA data, a null model with no covariates is required in order to investigate how much variance in the dependent variable i.e. reading achievement can be attributed to within and between clusters (SC).

Second, given the explorative nature of the analysis this study invokes, this study adopted the “top-down” approach introduced by (SC) to build the model, this approach comprise two steps as described by (SC): first, including all the fixed effects in the model then removing insignificant effect(s), second, including all random effects then removing significant ones.

Third, to further improve the model, modification indices (MIs) (SC) were used to diagnose the model fit and to identify its misspecification. MIs reveal the decrease in the χ^2 statistics of the model with 1 df if a given parameter is freed from model constraints. Despite the fact that there is no certain strict cut-off for how large the MIs should be to be considered, a 3.84 decrease in χ^2 with 1 df indicates a $P = 0.05$ fit improvement in the model (SC, p. 23), a significant fit this study adopted. Parameters with high MIs were freed one at a time, in descending order, as each freeing will results in change in other parameters (SC).

Stick to MIs, are the expected parameter change (EPC), which are measures that estimates the expected change in the value of a certain parameter if that parameter was freed (SC). (hou and Bentler, 1993) developed fully—in contrast to Kaplan 1989 partially— standardized EPC which is, unlike Kaplan's, invariant under different scalings of observed and latent variables, hence favored by this study. Mplus provides both MIs and fully standardized EPC for all fixed and constrained parameters.

Results that have unrealistic values e.g. negative, was not considered. Last but not least, it is important to highlight that the modification process was not blindly decided upon statistics only, but thoughtfully justified through theoretical basis. When evaluating the SEM goodness-of-fit, this study referred to the widely used (Hu & Bentler, 1999), cut-offs that is ≤.06 for root-mean-square error of approximation (RMSEA); ≤.08 for standardized root mean square residual (SRMR); ≥.95 for comparative fit index (CFI) and Tucker-Lewis index

(TLI). These cut-offs can be less strict when evaluating the models with WLSMV estimator, that are FA models (SC).