JPII 12 (1) (2023) 96-111

Jurnal Pendidikan IPA Indonesia

http://journal.unnes.ac.id/index.php/jpii

USING MULTILEVEL MODELLING TO EVALUATE SCIENCE LITERACY AND TECHNOLOGY COURSE OF THE INDONESIAN NON-SCIENCE STUDENTS

B. Setiaji*[1], P. H. Santoso[2,6], K. N. Aziz[3], H. Retnawati[4], M. Khairudin[5]

[1,3]Department of Physics Education, Universitas Negeri Yogyakarta, Indonesia
[2,4,5]Graduate School of Educational Research and Evaluation, Universitas Negeri Yogyakarta, Indonesia
[6]Department of Physics Education, Universitas Sulawesi Barat, Indonesia

DOI: 10.15294/jpii.v12i1.41457

Accepted: December 28th, 2022. Approved: March 30th, 2023. Published: March 31st, 2023

ABSTRACT

In this paper, a science literacy and technology course has been designed and implemented to strengthen the national initiative empowering scientifically literate Indonesian society. This paper is intended to evaluate to what degree non-science undergraduate students can perform this course. The diverse background of non-science students who participated in this study led to the challenge to evaluate their performance more comprehensively contemplating the nested structure of students' department and faculty setting. In light of the hierarchical nature of the student data, multilevel modelling was used to conduct the analysis. The first level of analysis involved students' performance and affective attributes measured using demonstrated science literacy assessment (SLA-D) and motivational beliefs (SLA-MB) respectively. Then, the subsequent level of analysis comprised demographic factors gathered from the institutional record. Findings demonstrated that the impact of demographic factors on the students' performance of science literacy was not substantial. Different settings of students' department and faculty level drove the association between affective factors and the learning process toward science literacy courses substantially. The multilevel approach controlled the equitable student assessment within the nature of students' data structure. This paper suggests an implication of advancement regarding educational data analysis and examines the effectiveness of science literacy courses for higher institutions specifically for non-science majors.

Keywords: multilevel modelling; non-science students; science literacy

INTRODUCTION

Many research works have been attempted to develop scientific literacy worldwide by the science education research community (Impey et al., 2011; Queiruga-Dios et al., 2020; Odden et al., 2021; Santoso et al., 2022). Survey reports established by the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) undoubtedly impact educational initiatives worldwide, including the Indonesian science education system (Cheema, 2017; Rachman et al., 2021; Ustun et al., 2022). As revealed by those survey programs, unsatisfying Indonesian students' performance forces the government to revise its curriculum, echoing this challenge persuasively. Since 2020, the Indonesian Ministry of Education, Culture, Research, and Technology (KEMDIKBUDRISTEK) has piloted and disseminated curricular innovation in terms of 'MERDEKA BELAJAR' to shift the former national assessment program for the foci of scientific literacy education (Nurjati et al., 2022). After that, the newer Indonesian assessment agenda is set to harness secondary school students regarding

*Correspondence Address
E-mail: bayu.setiaji@uny.ac.id

the science literacy domain as measured by the recent dimension of the PISA and TIMSS survey (Salamah et al., 2022). Without neglecting the essence of content knowledge delivered by other disciplines, this breakthrough is planned to propagate better student performance than the past international science survey administered by reputable organizations like PISA and TIMSS. At the same time, we are confident this policy can be progressive and must invite the stakeholders and policymakers engaged in building the sustainability of the future Indonesian civilisation including higher education as endeavoured by this paper.

Higher education institutions must be a groundbreaker to take part in the area of education and research expertise. Universitas Negeri Yogyakarta (UNY) is one of the Indonesian teacher education institutions (TEIs) including science education for secondary school students of the whole nation. Recently, UNY has focused more on strengthening the present national vision of fostering science literacy. At UNY, the science literacy and technology course (MKU 6217) has been programmed since 2020 as compulsory according to the syllabus of the first-year undergraduate students across the department and faculty (BAKK UNY, 2019). This course is predominantly offered for non-science students to furnish their competencies in being future Indonesian educators. Albeit they will be likely to fall into non-science careers, they are prepared with several aspects of the nature of science encompassing the principles of scientific thinking. Essentially, they are anticipated to be set up as the imminent generation of society after they proceed from their undergraduate education. Meanwhile, they must be still projected to be prospective teachers professionally in each field but are equipped to empower scientifically literate Indonesian society.

This course, as previously highlighted, is crucial, particularly for TEIs like UNY in educating the prospective teacher for the Indonesian community. UNY has just started the course implementation in recent years. It would be the appropriate moment to evaluate the implementation of the science literacy program for course enhancement. Admittedly, effective science literacy instruction requires the attention of evaluation during its implementation (Hobson, 2003; Hastuti et al., 2020). In this study, students' data have been harvested from four different non-science majors in the 2022/2023 academic year, namely marketing, accounting, Javanese language, and dance education. Due to the distinct majors, students participating in this study evidently generate diverse backgrounds, prior knowledge, family resources, and other variables related to them (Creswell, 2015). Then, students' performance throughout the course must be diverse and mixed. Consequently, it would be challenging to better understand the course's impact on the students' science literacy. Within this context, our study critically needs a more thorough approach in evaluating to what degree our non-science students have performed science literacy course.

Relevant works can be cited to understand state of the art regarding how researchers have designed and evaluated science literacy instruction specifically in the context of higher education and more specifically for non-science majors as focused by our study. Our research is unique and challenging with the argument of investigation of science literacy courses for non-science college students rooted in institutionally diverse backgrounds. A plethora of methods has been approached in evaluating science literacy courses in an interdisciplinary context of college education (Efthimiou & Llewellyn, 2004; Parkinson & Adendorff, 2004; Hobson, 2008; Impey, 2013; Ross et al., 2013; Sjöström & Eilks, 2018; Surpless et al., 2018; Hamper & Meisel, 2020). From the mentioned literature, we can simply summarize that few researchers consider the nested impact that emerged from the nature of students' data. In fact, students' data is mostly situated within the hierarchy of college administrative areas. Thus, clustered context must be faced by our students and has the potential influence for rigor analysis. Underestimating this nature will contribute to students' performance evaluation bias that should be carefully prevented.

Broadly speaking, students' performance is one proxy of the educational process that is commonly reported by education scholars. Many analytical approaches have been proposed to better measure, examine, and evaluate students' performance during the learning processes (Ding, 2019). Dealing with the issue of hierarchical data above, multilevel modelling, rooted in regression analysis, is then proposed further to consider the nested structure of students' learning (Finch et al., 2016). In this study, the multilevel modelling approach will enrich our understanding of students' performance in science literacy courses and the corresponding association with some latent factors such as affective attributes and demographic variables as frequently reported by scholars using PISA and TIMSS data (Mohammadpour et al., 2015; Ersan & Rodriguez, 2020; You et al., 2021; Ustun et al., 2022).

Utilising multilevel modelling for data analysis requires one to take the clustered level of data endorsed by data points into consideration. In the context of this study, our first level of data points is students. They can generate educational data and associated factors that will bring potential influences on the students' learning. Adjacent to the cognitive aspect, students can possess affective attributes throughout their learning process. Also, this embodies the first level of our data since it is created by students. Afterwards, the departmental factors would contribute to the subsequent hierarchy of our students' data. It can be understood based on demographic variables recorded by the institutional information systems. Students are managed in certain departments and faculty. Demographic variables created by the nested structure of students' departments and faculty can correlate with the potential difference in students' performance (Salehi et al., 2019; Kanim & Cid, 2020; Simmons & Heckler, 2020). Therefore, multilevel modelling must be suitable to make the analysis more comprehensive.

Prior works have extensively documented the association of students' performance toward affective attributes (Fives et al., 2014; Bellová et al., 2021; Rudolph, 2020; Fortus et al., 2022) and demographic variable (You et al., 2021; Ustun et al., 2022) within scientific literacy learning. They have argued that affective results can substantially influence students' scientific literacy. Nevertheless, several researchers approach distinct construction of the affective measure. Thus, the implication of their results to the other contexts needs to be further examined. For instance, Fortus et al. (2022) describes the definition of the affective attribute into four constructs, namely interest, attitude, self-efficacy or self-concept, and motivation. Another idea has been proposed by Fives et al. (2014) that have developed the motivational belief measurement for their science literacy assessment (SLA).

Furthermore, demographic variables may contribute to the variance of the scientific literacy assessment as reported by Ustun et al. (2022) and You et al. (2021) in the context of PISA data. They discover that scientific literacy can be substantially influenced by demographic variables such as economic/ social/ cultural status (ESCS). Nevertheless, several scholars have shown that the demographic effects are mixed and still inconclusive (Simmons & Heckler, 2020). There must be other more important factors to make the investigation more generalizable.

To complete the missing area of the prior works, this study is framed as an evaluation attempt toward the recent implementation of a science literacy and technology course designed by the UNY curriculum. The nested data structure of students' performance, affective attributes, and demographic variables are measured and analysed using multilevel modelling techniques within three levels of data (student, department, faculty). The following two research questions were proposed to guide this study: RQ 1: To what degree does the mean difference of students' performance and affective attribute on science literacy courses vary within department level, faculty level, and demographic variables?; RQ 2: To what degree does the department and the faculty levels vary the dependence between students' performance on science literacy course, affective attribute, and demographic variable?

Investigating the influential factors toward students' learning would inform scholars, educators, and practitioners in designing science learning more equitable. This study evaluated the recent implementation of science literacy and technology course based on data harvested from the 2022/2023 academic year. The analysis is conducted based on multilevel modelling results of the dependence between students' performance, affective attribute, and demographic variables clustered from the nature of college students' department and faculty setting. Evidence reported by this paper can be helpful for opening discussion rooms concerning educational data analysis dealing with the complex structure of students' data that must be warranted.

METHODS

What this paper aims to do is a quantitative survey using the approach of multilevel modelling to evaluate Indonesian non-science students' performance in science literacy courses associated with affective attribute and demographic variables hierarchically gathered from four distinct non-science departments and two different faculties involved in the study.

As briefly introduced in the preceding section, the current study was circumstanced within the science literacy and technology course (MKU 6217) administered by Universitas Negeri Yogyakarta (UNY) during the first term of the 2022/2023 academic year. This course was compulsory for non-science majors. Students were taught about the nature of science, the scientific method, the development of human civilization, the universe as a system, science technology and implementation, and culminated with the topic regarding big ideas in science. The syllabus of our science literacy course (BAKK UNY, 2019) was designed according to the university model and the aforementioned national movement. Due to the ongoing transition amid the pandemic situation, distance learning was still one of the learning modes. Moodle-based learning management system (LMS) developed by UNY prior to the pandemic, 'BESMART', was utilized in delivering, managing, and administering this course to the whole students (Priyambodo, 2016; Surjono et al., 2017). Figure 1 depicts the learning dashboard hosted by BESMART for our science literacy and technology course.

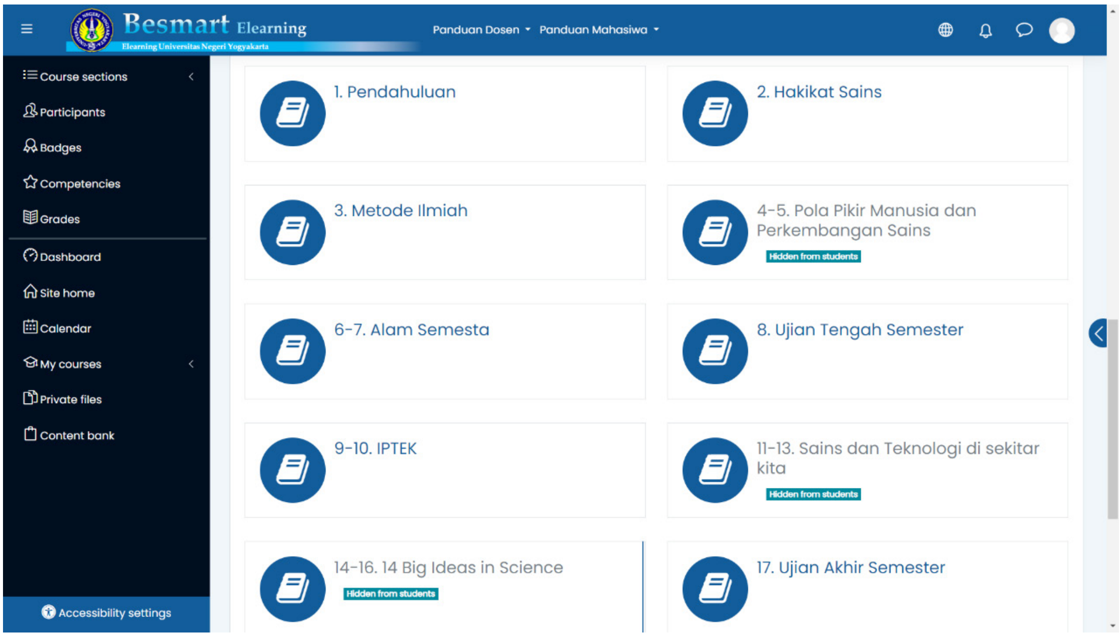

Figure 1. A Snippet of the BESMART Dashboard of Science Literacy and Technology Course

The first author (B.S.) and the third author (K.N.A.) were assigned as the lecturer of this course across four non-science undergraduate departments under two faculties. The participant of this study was determined on a voluntary basis without neglecting the representativeness of non-science students enrolled at UNY. They were recruited from the department of marketing (n = 38), accounting (n = 39), Javanese language (n = 51), and dance education (n = 50). Department of marketing and accounting education were registered under the administration of the Faculty of Economics (FE) and two remaining study programs were administered under the Faculty of Language and Arts (FBS). In total, 178 students participated in this study. Yet, those students which were recorded as missing some data points either belong to one or some variables had been scheduled for the second attempt of the survey session. Nevertheless, eighteen students should be deleted for further data analysis due to their lack of awareness to our provided second chance.

The first level of students' data was gathered using multiple points of in-class assessment including students' performance in the science literacy course, midterm exam, and their affective attributes. Students' performance then would be our dependent variable. At the same level, the midterm exam and affective attribute were simultaneously measured. It would be the predictors that might influence students' performance. Subsequently, each department and faculty share the parallel characteristics corresponding to the demographic variables that emerged from the students' data. In this study, the demographic data were compiled from the institutional record. The department was the second level of data, and subsequently, the faculty was the third level of data.

A measure of scientific literacy assessment (SLA) disseminated by Fives et al. (2014) was employed to probe students' performance. SLA covered the 'demonstrated' (SLA-D) and the 'motivational belief' scientific literacy assessment (SLA-MB). Twenty-six multiple-choice items of SLA-D were utilized as a proxy of students' performance in this course. SLA-MB was employed in our survey of affective attributes. Obviously, the English version of the SLA should be translated and adjusted to the context of the Indonesian community. Content validity of two experts with teaching and research experience in science literacy education for more than ten years had been conducted. They were the fourth (H.R.) and fifth author (M.K.) of this paper.

As a formative assessment, five open-ended items were also administered by the midterm examination. We tested students in several aspects of science literacy and were also inspired from constructs explained by Fives et al. (2014) but in other forms of open questions rather than dichotomous responses as measured by SLA-D above. Some evidence suggests that science literacy should be measured using an open-ended format rather than a closed response (Miller, 1998). Moreover, the grading rubric of the midterm exam was developed by the lecturer of this course (the first and third author of this paper). Face validity had been conducted to those experts parallel to the former validation step of the Indonesian version of SLA-D and SLA-MB.

Theoretically, three constructs were demonstrated by SLA-MB underlying the affective attributes of science literacy. They were comprised of value of science (VOS), what can I do in science (DIS), and what I believe about science (BAS) (Fives et al. 2014). Twenty-five observed items of a five-point Likert scale were distributed under these three factors. Using the students' response data (n = 160), we discovered plausible reliability as discovered by Fives et al's result (α_{alll}=0.781, α_{VOS}=0.868, α_{DIS}=0.786, α_{BIS}=0.903). Exploratory factor analysis (EFA) was also used to validate the alignment of Fives et al's construct toward the students' state of representation. We discovered three well-defined and unique components. The eigenvalues of SLA-MB achieved more than unity. Therefore, it was consistent with prior findings (Fives et al. 2014) that three constructs of SLA-MB can be used independently of one another.

The demographic variables were collected from the institutional registrar record. In this study, ten demographic variables were harvested and examined for further analysis. They were gender (1 = male, 2 = female), admission pathway (1 = 'SNMPTN' or national based college admission system via portfolio, 2 = 'SBMPTN' or national based college admission system via written test, 3 = 'SM Prestasi' or on-campus based college admission system via portfolio, 4 = 'SM Utul' or on-campus based admission system via written test), tuition funding (1 = subsidized, 2 = non-subsidized), scholarship holder (1 = yes, 2 = no), high school background (1 = science, 2 = non-science), residence (1 = North Sumatera, 2 = West Sumatera, 3 = Riau, 4 = Jambi, 5 = Bengkulu, 6 = Bangka Belitung, 7 = Jakarta, 8 = West Java, 9 = Central Java, 10 = Yogyakarta, 11 = East Java, 12 = East Kalimantan), father/ mother education (1 = master, 2 = bachelor, 3 = diploma, 4 = senior high school, 5 = junior high school, 6 = elementary school, 7 = uneducated), and father/ mother monthly income (1 = more than IDR4.000.000, 2 = IDR3.000.000–4.000.000, 3 = IDR2.000.000–3.000.000, 4 = IDR1.000.000–2.000.000, 5 = less than IDR1.000.000). Table 1 describes the summary of the number (n) and proportion (%) on a class of each attribute.

Table 1. Participants Based on the Demographic Attributes within the Departmental Distribution

Attribute	Class	Faculty of Economics				Faculty of Language and Arts				Total	
		Marketing		Accounting		Javanese		Dance			
		n	%	n	%	n	%	n	%	n	%
Gender	Male	5	.03	12	.08	18	.11	4	.03	39	.24
	Female	**33**	**.21**	**24**	**.15**	**29**	**.18**	**35**	**.22**	**121**	**.76**
Admission pathway	SNMPTN	0	0	0	0	13	.08	0	0	13	.08
	SBMPTN	0	0	0	0	14	.09	**14**	**.09**	28	.18
	SM Prestasi	**35**	**.22**	**24**	**.15**	5	.03	12	.08	**76**	**.48**
	SM Utul	3	.02	12	.08	**15**	**.09**	13	.08	43	.27
Funding	Subsidized	0	0	0	0	**27**	**.17**	14	.09	41	.26
	Non-subsidized	**38**	**.24**	**36**	**.23**	20	.13	**25**	**.16**	**119**	**.74**
Scholarship	Yes	8	.05	1	.01	6	.04	10	.06	25	.16
	No	**30**	**.19**	**35**	**.22**	**41**	**.26**	**29**	**.18**	**135**	**.84**

Attribute	Class	Faculty of Economics				Faculty of Language and Arts				Total	
		Marketing		Accounting		Javanese		Dance			
		n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
High school Major	Science	9	.06	12	.08	15	.09	17	.11	53	.33
	Non-science	**29**	**.18**	**24**	**.15**	**32**	**.20**	**22**	**.14**	**107**	**.67**
Residence	North Sumatera	0	0	2	.01	0	0	0	0	2	.01
	West Sumatera	0	0	0	0	0	0	1	.01	1	.01
	Riau	0	0	0	0	0	0	1	.01	1	.01
	Jambi	1	.01	0	0	0	0	0	0	1	.01
	Bengkulu	0	0	0	0	0	0	1	.01	1	.01
	Jakarta	0	0	2	.01	1	.01	0	0	3	.02
	West Java	1	.01	5	.03	0	0	4	.03	10	.06
	Central Java	**23**	**.14**	13	.08	**20**	**.13**	7	.04	**63**	**.39**
	Yogyakarta	10	.06	**14**	**.09**	17	.11	**14**	**.09**	55	.34
	East Java	3	.02	0	0	19	.12	9	.06	21	.13
	East Kalimantan	0	0	0	0	0	0	1	.01	1	.01
	West Nusa Tenggara	0	0	0	0	0	0	1	.01	1	.01
Father education	Master	0	0	0	0	0	0	1	.01	1	.01
	Bachelor	11	.07	4	.03	11	.07	10	.06	36	.23
	Diploma	3	.02	0	0	2	.01	1	.01	6	.04
	Senior	**18**	**.11**	**25**	**.16**	**13**	**.08**	**14**	**.09**	**70**	**.44**
	Junior	0	0	3	.02	10	.06	7	.04	20	.13
	Elementary	5	.03	4	.03	9	.06	5	.03	23	.14
	No school	1	.01	0	0	2	.01	1	.01	4	.03
Mother education	Master	2	.01	0	0	0	0	0	0	2	.01
	Bachelor	9	.06	6	.04	4	.03	9	.06	28	.18
	Diploma	4	.03	3	.02	3	.02	2	.01	12	.08
	Senior	**16**	**.10**	**13**	**.08**	**21**	**.13**	**18**	**.11**	**68**	**.43**
	Junior	3	.02	9	.06	14	.09	4	.03	30	.19
	Elementary	4	.03	5	.03	4	.03	4	.03	17	.11
	No school	0	0	0	0	1	.01	2	.01	3	.02
Father income	> 4 mio	7	.04	7	.04	2	.01	4	.03	20	.13
	3–4 mio	2	.01	2	.01	7	.04	1	.01	12	.08
	2–3 mio	9	.06	6	.04	5	.03	6	.04	26	.16
	1–2 mio	**15**	**.09**	**12**	**.08**	**17**	**.11**	13	.08	**57**	**.36**
	< 1 mio	5	.03	9	.06	16	0.1	**15**	**.09**	45	.28
Mother income	> 4 mio	3	.02	1	.01	1	.01	3	.02	8	.05
	3–4 mio	7	.04	1	.01	4	.03	2	.01	14	.09
	2–3 mio	3	.02	4	.03	2	.01	2	.01	11	.07
	1–2 mio	6	.04	4	.03	4	.03	6	.04	20	.13
	< 1 mio	**19**	**.12**	26	.16	36	.23	26	.16	107	.67

Note: The majority class of each demographic attribute is indicated in bold

In Table 1, we could immediately discover that the majority of our students were female (76%), admitted by the university through on-campus admission via portfolio or 'SM Prestasi' (48%), non-subsidized funding (74%), no scholarship (84%), non-science high school background (67%), Central Java people (39%), live with senior high school graduated father (44%) and mother (43%), and economic status with father's and mother's monthly income between IDR 1 – 2 million and less than IDR 1 million respectively. In conclusion, we collected three variables from in-class assessments and ten variables extracted from university registrar records. For the subsequent explanation, Table 2 summarises the description, code (for the equation explanation below), measurement tool, and data type endorsed by each corresponding variable of each variable gathered by this study.

Table 2. Summary of the Investigated Variables Compiled by this Study

No	Variable	Code	Source	Scale	Type
1	Students' performance	SPerf	26 multiple-choice items of SLA-D	100-point	Continuous/ interval
2	Midterm exam	MidTerm	5 open-ended items	100-point	Continuous/ interval
3	Affective attribute	Aff	25 questionnaire items of SLA-MB	5-point Likert scale	Categorical/ ordinal
4	Gender	Gend	Institution registrar record	2-point	Nominal
5	Admission pathway	Adm		4-point	Nominal
6	Funding	Fund		2-point	Nominal
7	Scholarship	Sch		2-point	Nominal
8	High school Major	HS		2-point	Nominal
9	Residence	Res		12-point	Nominal
10	Father education	FEdu		7-point	Nominal
11	Mother education	MEdu		7-point	Nominal
12	Father income	FInc		5-point	Nominal
13	Mother income	MInc		5-point	Nominal

Students' performance (SPerf) in this study as measured by SLA-D was scored using the rubric of the correct option disseminated by Fives et al. (2014). Of 26 multiple-choice items, students' response was scored and transformed into 100 scale points. Then, before the affective scale as quantified by SLA-MB (Aff) was summed up, students' responses to the BAS factor should be reversed due to the negative items. The midterm exam (MidTerm) was also graded using the same scale as determined by the students' performance.

Prior to the multilevel modelling, analysis of variance (ANOVA) was employed in RQ1 to test the mean difference in students' performance, midterm exam, and affective attribute among the class of each department, faculty, and demographic aspect. ANOVA was calculated to justify that department and faculty levels can influence the variance of students' performance. As well, ANOVA would tell us about the potential demographic factors that should be included in the equation of the multilevel model (RQ2). Merely significant differences ($\alpha<0.05$) among the class on each factor that would be included in the multilevel equations influencing the students' performance on science literacy as the target variable.

In RQ2, a multilevel modelling approach was utilized to deal with the nested data structure of the student's department and faculty level (Finch et al., 2016). A multilevel modelling technique would be fit to process the clustered students' data within four distinct departments and under two distinct faculties. Two-level modelling was intentionally first analyzed. It followed the three-level modelling involving the faculty level in the subsequent. Figure 2 shows the data structure with three levels in which *i*, *j*, and *k* represent student, department, and faculty levels respectively.

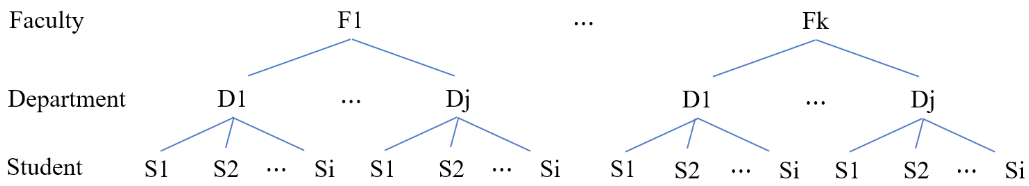

Figure 2. Network Depicting A Data Structure of Student (Level 1), Department (Level 2), and Faculty (Level 3)

The first step of the analysis was building the null model or the baseline in which none of the predictors were included in the equation. The null model was used as a baseline for model building and subsequent comparison. The null model was formulated as follows.

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \varepsilon_{ij} \quad (1)$$

$$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \varepsilon_{ijk} \quad (2)$$

As a reference, we adapted the mathematical notation of multilevel modelling from Finch et al. (2016). Equation (1) refers to the two-level modelling and equation (2) accounts for the three-level modelling consecutively. SPerf_{ij} refers to the students' performance of the *i*-th students under the *j*-th department, γ_{00} refers to the grand intercept mean of the *j*-th department, U_{0j} accounts for the random effect of the *j*-th department, and ε_{ij} term indicates the student-level random error that is not explained by the model. Equation (2) is a bit similar to the former. Yet, this is built for the higher level thus we find a *k* subscript indicating the level of students' faculty. As γ_{00} above, δ_{000} is the grand intercept mean of the *j*-th department nested in the *k*-th faculty. Accordingly, V_{00k} represents the grand intercept mean of the *k*-th faculty. Eventually, U_{0jk} refers to the random effect and ε_{ijk} is the random error that is unable to be explained by the model.

After that, the subsequent models were made based on the initial finding of RQ1 using ANOVA and complemented with intra-class correlation (ICC) results based on the null model. Multilevel modelling was fitted using the 'lmer' function of the 'lme4' package of the R programming environment (Bates et al., 2015). The restricted maximum likelihood (REML) estimation approach was selected since it has proven more accurate than maximum likelihood estimation (MLE) for estimating variance parameters (Luo et al., 2021). Overall, those built models were compared using the Akaike information criterion (AIC) (Matuschek et al., 2017). As a rule of thumb, the best model should be interpreted based on the lowest AIC of the model comparison.

RESULTS & DISCUSSION

Initial information from RQ1 justified potential association within features discovered based on the statistical analysis. ANOVA will test the mean difference of certain variables among clusters or groups. In this study, the cluster must be department, faculty, and demographic variables as described specifically in Table 2. Then, ICC is a measure to what degree categorical variables such as demographic factors can correlate to certain variables, for instance, students' performance, midterm exam, and affective attributes. Therefore, a significant result from ANOVA and a plausible ICC value of a predictor would be the indication that those significant variables should be taken into account in the multilevel model. The results of RQ1 are given in Table 3.

First, we highlighted a significant mean difference in students' science literacy (SPerf) and midterm exam (MidTerm) among the department and the faculty level ($p < 0.05$). Furthermore, there was a plausible correlation between department and faculty level toward students' performance on science literacy (SPerf). The midterm exam also correlated with the department level, yet we discovered no correlation with the faculty level. Those results indicated that students' performance in science literacy courses would be empirically influenced by differences between the department and the faculty. Hence, one can justify equations (3) and (4) which add the midterm exam (MidTerm) to the model.

Table 3. ANOVA Results of the Mean Difference of the Student's Performance (Sperf), Affective Attribute (Aff), And Midterm Exam (Midterm) Among the Class on Each Student's Department, Faculty, and Demographic Factor

Variable	SPerf			Aff			MidTerm		
	F	p	ICC	F	p	ICC	F	p	ICC
Department	11.21	**0.000**	0.209	0.162	0.922	0.000	6.902	**0.000**	0.129
Faculty	19.04	**0.000**	0.185	0.002	0.964	0.000	0.130	0.719	0.000
Gender	0.116	0.734	0.000	0.000	0.988	0.000	1.405	0.238	0.007
Admission Pathway	1.968	0.121	0.026	1.360	0.257	0.000	0.719	0.542	0.000
Funding	0.617	0.433	0.000	0.573	0.450	0.000	0.265	0.607	0.000
Scholarship Holder	0.600	0.440	0.000	0.008	0.929	0.000	0.010	0.920	0.000
High School Major	0.746	0.389	0.000	5.182	**0.024**	0.056	0.945	0.333	0.000
Residence	1.056	0.401	0.004	0.888	0.553	0.000	0.707	0.730	0.011
Father Education	0.867	0.521	0.000	1.329	0.247	0.030	0.613	0.720	0.000
Mother Education	1.162	0.330	0.000	2.245	**0.042**	0.055	0.228	0.967	0.000
Father Income	0.793	0.532	0.000	0.557	0.694	0.000	0.619	0.650	0.000
Mother Income	0.511	0.728	0.000	1.185	0.320	0.020	1.059	0.379	0.000

Note: The significant mean differences (α<0.05) among the class of each student's department, faculty, and demographic aspect are indicated in bold.

$$\text{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\text{MidTerm}_{ij} + \varepsilon_{ij} \quad (3)$$

$$\text{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \delta_{100}\text{MidTerm}_{ijk} + \varepsilon_{ijk} \quad (4)$$

Models 3 and 4 added the γ_{10} and δ_{100} term. They enumerate the regression coefficient between its predictor and the outcome variable. Second, we find no significant mean difference in students' affective attributes among the department and faculty levels. Conversely, Fives et al. (2014) argued that there is a strong correlation between affective measures and students' science literacy. Indeed, Figure 3 could be consulted to support Fives et al. (2014) which depicts a scatter plot matrix of SPerf, Aff, and MidTerm Pearson correlation values.

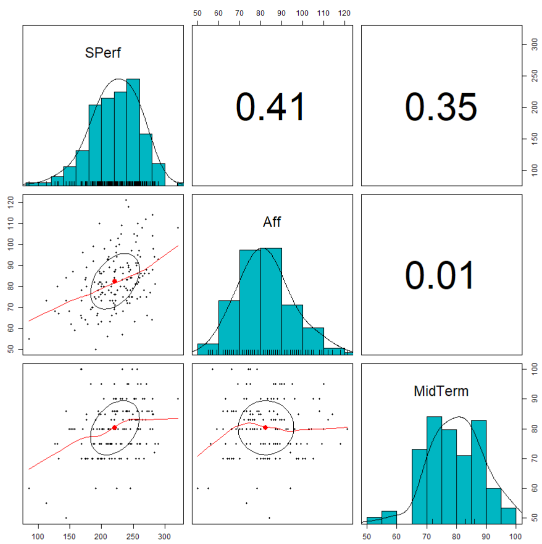

Figure 3. Scatter plot and Pearson Correlation between SPerf, Aff, and MidTerm. Association between SPerf toward both Aff and MidTerm are Described (r = 0.41 and r = 0.35 respectively). The Independence within Predictors (Aff and MidTerm) is also Visualized (r = 0.01).

One can see that there is a significant correlation between SPerf and Aff. Therefore, the affective attribute could not be neglected to understand its association with students' science literacy. Thus, the third model added the affective attribute as formulated in equations (5) and (6) as follows.

$$\mathrm{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\mathrm{MidTerm}_{ij} + \gamma_{20}\mathrm{Aff}_{ij} + \varepsilon_{ij} \quad (5)$$

$$\mathrm{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk} + \delta_{100}\mathrm{MidTerm}_{ijk} + \delta_{200}\mathrm{Aff}_{ijk} + \varepsilon_{ijk} \quad (6)$$

Clearly, γ_{20} and δ_{200} accounted for the coefficient of the affective attributes toward the students' science literacy.

Third, there is a significant mean difference ($p < 0.05$) of affective measure among the class of students' high school background and social status in terms of their mother's education. Surprisingly, there was no significant impact of other demographic variables such as gender, admission pathway, funding, scholarship holder, residence, father's education, and parents' income. Hence, those non-significant demographic factors should be omitted. The next model should invite those results and we added two significant demographic variables (HS and MEdu) as explained by equation (7) for the two-level and (8) for the third-level model as follow.

$$\mathrm{SPerf}_{ij} = \gamma_{00} + U_{0j} + \gamma_{10}\mathrm{MidTerm}_{ij} + \gamma_{20}\mathrm{Aff}_{ij} + \gamma_{30}\mathrm{HS}_{ij} + \gamma_{40}\mathrm{MEdu}_{ij} + \varepsilon_{ij} \quad (7)$$

$$\mathrm{SPerf}_{ijk} = \delta_{000} + V_{00k} + U_{0jk}\delta_{100}\mathrm{MidTerm}_{ijk} + \delta_{200}\mathrm{Aff}_{ijk} + \delta_{300}\mathrm{HS}_{ij} + \delta_{400}\mathrm{MEdu}_{ij} + \varepsilon_{ijk} \quad (8)$$

Where γ_{30} and δ_{300} indicates the impact of different students' high school background on their students' performance in science literacy course. Accordingly, and corresponds to the dependence of students' performance on science literacy with the status of mother education.

The current presentation will demonstrate our multilevel modelling results after the fitting of those eight equations. The results are given in Table 4. To make the description easier to interpret for the readers, we commence the two-level modelling results (models 1, 3, 5, and 7) that will be followed up with three-level findings (models 2, 4, 6, and 8).

Model 1 in the second column of Table 4 was built to examine the department-level association toward students' performance on science literacy. For the fixed effects, we discovered a significant intercept (γ_{00}, $p < 0.05$). The value within parentheses in Table 4 was the corresponding standard error. This significant intercept indicated that the department level could have the possibility to influence students' performance on science literacy that could be correlated with the random effects. The values reported by the section on random effects in Table 4 were the variance component and the corresponding standard deviation within parentheses. Intuitively, variance and standard deviation results could be interpreted as the extent to which the intercept (U_{0j}) varies by department level. Then, the residual component reported by Table 4 was the ε_{ij} term in equation (1).

Respectively, model 3 added the MidTerm variable into the equation. We also discovered significant results indicating that midterm examinations could influence the students' science literacy. In model 3, we also discovered a significant intercept (γ_{00}, $p < 0.05$) and coefficient of the midterm exam (γ_{10}, $p < 0.05$). Model 3 reported a greater standard error on its intercept than the null model. Conversely, we found a diminished pattern of random effects both in terms of the variance and the standard components of the intercept (U_{0j})) and residual (ε_{ij}). This indicated that multilevel models more precisely estimated the standard errors for our parameters.

In model 5, we added the affective attribute into the model. However, there was no significant intercept (γ_{00}, $p > 0.05$) yet it had a greater standard error than the previous models. The coefficient of the midterm exam (γ_{10}) was still significant ($p < 0.05$) with similar results and decreased slightly with model 3. An affective attribute (γ_{20}, $p < 0.05$) was discovered to significantly influence students' science literacy and this is consistent as reported by Fives et al. (2014). The diminishing pattern of intercept (U_{0i}) and residual (ε_{ij}) was also discovered formerly.

Regarding model 7, non-significant coefficients were discovered both in the student's high school background (γ_{30}, $p > 0.05$) and the mother education variable (γ_{40}, $p > 0.05$). Those variables thus cannot be concluded as influential factors to predict students' science literacy. This result inclusively differed from the previous literature mentioned in the introduction above (Mohammadpour et al., 2015; Ersan & Rodriguez, 2020; You et al., 2021; Ustun et al., 2022). The intercept of the fixed effects (γ_{00}, $p > 0.05$) was non-significant as reported by model 5 of affective impact.

Finally, the overall model has been reported in the last section of Table 4. These results can be helpful in characterizing and comparing two-level models. Degree of freedom (*df*), number of groups (*N*), and observations are reported to characterize the models. Among the four models, the lowest AIC was reported by model 5. Thus, model 5 was the best two-level model that fit the data analyzed in this study.

As in the two-level modelling, the null model was also built for the three-level model as the third column of Table 4. A significant intercept (δ_{00}, $p < 0.05$) was also discovered with a greater standard error than model 1 above. Therefore, incorporating the faculty level into the model captures the more accurate model to understand students' performance in science literacy. Then, the variance and standard deviation of the intercept from the department level (U_{0jk}) were lower than model 1 but with the same residual value (ε_{ijk}). This remaining value can be shared with the added information varied by the faculty level (V_{00k}).

Table 4. Multilevel Modelling Results

Parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Fixed effects								
Intercept	221.46*(9.545)	221.54*(13.18)	119.86*(26.47)	116.86*(27.85)	21.10(28.12)	18.26(29.44)	-4.00(36.36)	-8.96(37.49)
Midterm exam			1.27*(0.3132)	1.31*(0.31)	1.28*(0.28)	1.32*(0.27)	1.34*(0.27)	1.39*(0.27)
Affective					1.18*(0.18)	1.19*(0.18)	1.18*(0.19)	1.18*(0.19)
High school							-5.72(5.20)	-5.55(5.19)
Mother Edu							16.96(22.68)	16.84(22.59)
Random effects								
Department-level								
Intercept	332.7(18.24)	168.1(12.96)	248.9(15.78)	63.58(7.974)	240.7(15.51)	52.47(7.24)	238.5(15.44)	35.06(5.92)
Faculty-level								
Intercept		247.4(15.73)		274.31(16.56)		279.03(16.70)		299.21(17.30)
Residual	1257.3(35.46)	1257.2(35.46)	1150.9(33.92)	1150.75(33.92)	905.2(30.09)	905.06(30.08)	890.0(29.83)	889.90(29.83)
Overall model								
df	3	4	4	5	5	6	12	13
N (Groups)	Dep (4)	Fac (2), Dep(4)	Dep (4)	Fac (2), Dep(4)	Dep (4)	Fac (2), Dep(4)	Dep (4)	Fac (2), Dep(4)
N (Observations)	160	160	160	160	160	160	160	160
AIC	1611	1612.4	1596.68	1597.90	1560.00	1560.99	1564.06	1564.63

Note: * $p < 0.05$

The midterm examination was a significant factor in predicting students' science literacy course (δ_{100}, $p < 0.05$). It was consistent with the result reported by model 3 above. A greater standard error of significant intercept (δ_{000}, $p < 0.05$) was also reported. For the random effects, we also discovered the diminished pattern of the intercept from the department level (U_{0jk}) and the model residual (ε_{ijk}).

In terms of model 6, the affective attribute was combined into the model. Relevant to model 5, it discovered the non-significant intercept (δ_{000}, $p > 0.05$). Meanwhile, it still obtained a greater standard error than model 4. The coefficient of the midterm exam was significant (δ_{100}, $p < 0.05$). Then, an affective attribute was discovered significantly (δ_{200}, $p < 0.05$) influencing students' science literacy and this is consistent as reported by model 5 in the two-level modelling result.

Regarding model 8, non-significant coefficients were discovered both in the high school variable (δ_{300}, $p > 0.05$) and the mother education variable (δ_{400}, $p > 0.05$). Hence, those variables could not be interpreted as influential factors to predict students' science literacy as reported by model 7 for two-level modelling results. Eventually, the lowest AIC among four three-level modelling results was reported by model 6 which merely considered students' level variable (midterm exam and affective attribute) into the model as performed by model 5 above. Therefore, model 6 best fit the three-level model based on the data analysed in this study.

This study is proposed to answer two research questions. In RQ1, we investigate to what extent the mean difference in students' science literacy and associated factors can be varied by different department levels, faculty levels, and demographic variables. In RQ2, we study to what degree the discovered department and faculty effects in RQ1 contribute to investigating the association between affective attributes and demographic variables toward students' performance in science literacy courses. In the next paragraphs, findings reported by this paper are discussed with discoveries of prior literature and further attention to solving limitations that can be driven by our selection of context and research procedures.

We can highlight five novel findings for the answer to RQ1. First, students' department and faculty levels have been evident as influential factors in making significant differences in their science literacy. This is consistent with prior studies that have been introduced earlier (Salehi et al., 2019; Kanim & Cid, 2020; Simmons & Heckler, 2020). Educational settings in which students are immersed during the learning process will construct their climate of learning. There are many factors incorporating the different contexts of students' learning process. It can be stimulated by complex factors regarding teachers' quality, class size, peer motivation, physical facility, and other difficult factors to identify. Therefore, controlling the measurement of students' performance using hierarchical and multimodal data should be worth maintaining for the rigorous assessment method.

Second, the formative assessment opted in the learning process can be a substantial predictor in evaluating the peak of the student's performance in science literacy courses. This is relevant to what has been emphasised by Hastuti et al. (2020) and Hobson (2003). They suggest that effective science literacy instruction requires ongoing attention to evaluation during its implementation. Formative assessment including midterm examination is one of the assessment points in controlling students during the learning process. Maintaining learning intention until the last part can predict students' success in learning science literacy in this study. This result may not come as a surprise since it may be characterized as common pedagogical knowledge for the majority of educators throughout their professional growth.

Third, supplementing the influence of formative assessment directly, a significant affective impact on students' science literacy has been confirmed by our study. It is immediately consistent with prior works that have been introduced earlier (Fives et al., 2014; Rudolph, 2020; Bellová et al., 2021; Fortus et al., 2022). Henceforth, it can be witnessed for science literacy educators that quoting students' attitudinal aspects in evaluating their performance must be imperative to create an assessment more responsive to the student's behaviour, attitude, and motivation. Adjusting course evaluation regarding those affective factors is the principle of authentic assessment of science education (Ratini et al., 2018; Nurjati et al., 2022, Salamah et al., 2022; Rofieq & Fauzi, 2022). Effective science literacy courses should be more sensitive to elicit this factor for their class assessment criteria. Recent predictive modelling studies by Mahmudah et al. (2022, 2020) discover that the importance of affective attributes can be effective in studying, predicting, and monitoring students' resilience during the learning process including science literacy education as the foci of science education studies recently (Sumarti et al., 2018; Romero et al., 2020; Widiana et al., 2020). It is inevitably a personal trait that students are commonly visualised as a social construct within the educational system. Nevertheless, many dimensions underpinned by the affective attribute can be more compelling for further attention in the psychoeducational assessment field.

Then, of ten demographic variables collected in this study, we just discover two variables that make substantial mean differences in students' science literacy. Surprisingly, gender, which is mostly predicted to greatly impact science learning in prior documentations (Cheema, 2017; You et al., 2021; Ustun et al., 2022), gives an unsubstantial factor to the variance of students' science literacy in this study. Gender bias remains a long debate within science education literature. Certain studies report that females will be beneficiaries through science education and performance is significantly distinct among students' gender (Ramdani et al., 2021; Susongko et al., 2021). Due to the limitation of the research focus of this paper, interested readers would be offered to study this in a more depth analysis using multiple methods quantitatively, qualitatively, or combined.

Last but not least, demographic discrepancies triggered by high school preparation and mother education have been evident to contribute to the attainment of science literacy indicated by the non-science undergraduate students. Despite the sphere of non-science culture dominating the students' environment, a diverse form of high school majors is already described in Table 1. Some students are educated with science majors in their high school history. This would likely influence the mental model of science literacy delivered to the students. Many studies demonstrate that school situations experienced by students could be one form of social-economic factor that should be further examined (Briones et al., 2022). In this study, the significant impact of high school majors will likely be invited to the multi-level paradigm. Moreover, it is surprising that parent contribution as mother education can be a potential factor toward what degree students can be scientifically literate. It is reasonable that the higher parents get an education, the more students are more likely to obtain opportunities and support (physically or mentally) during the educational process (Marzulina et al., 2018; Masud et al., 2019; Akram & Pervaiz, 2020; Güre et al., 2020; Kamba et al., 2020). Meanwhile, more studies must be warranted for this result can be described more extensively. Thus, it can be more generalisable for the wider population.

Furthermore, in RQ2, we can discuss two main findings elaborated for the answer of RQ2 in the context of multilevel modelling results. First, we can summarize the best multilevel model and the pattern of the estimation results demonstrated by two-level modelling in Table 4 above. Both two-modelling and three-modelling results decide that adding the midterm exam and affective variable is the most representative model of students' science literacy assessment in this study. Adjacent to the cognitive aspect, affective attribute is the common procedure in doing authentic assessment as recommended by the literature on science literacy education described earlier (Nurjati et al., 2022, Salamah et al., 2022). This is obviously in line with the third finding of RQ1 discussed formerly. On the other hand, in model 5, a significant intercept of the random effects driven by department and faculty setting is absent. This can be translated as the controller of equitable assessment criteria for students' evaluation. Nevertheless, inviting affective measures to consider the assessment aspect can be carefully constructed even though its latent factor must be admittedly difficult to measure.

Second, we discover the diminished pattern toward the standard error of intercept after controlling the model using the multilevel analysis. It can be understood that variance source from department and faculty level precisely influences the dependent variable and we should consider the improved analysis. Precisely, it is evident that hierarchical-based analysis can capture the students' science literacy more accurately as recommended by the prior study (Mohammadpour et al., 2015; Ersan & Rodriguez, 2020). The hierarchical difference during the learning process is the consequence of the school environment. Intuitively, this result can be a recommendation to the policymakers to ensure equity in education and facilitate a better educational infrastructure for all students. Balancing the quality among the schools is admittedly able to make students feel more supported. Thus, their performance could be boosted, and the educational initiatives by the nation would be realized.

Nevertheless, we believe this study may be driven by several sources of uncertainty caused by three constraints. First, errors and uncertainties could be introduced by the data collection processes and the limited non-science students involved in our study. This study is framed as an evaluation attempt toward implementing a science literacy course designed at UNY. Thus, the findings reported by this paper can be distinct in the case beyond other universities owned by the Indonesian education system and overseas.

Second, the affective measure is a latent construct. Many studies have reported diverse ideas to frame the definition of affective measure toward science literacy. Arguably, the selection of employing SLA-MB from Fives et al. (2016) must be further examined. Then, the selection of SLA items administered by this study could be extensively developed particularly for measuring the diverse form of scientific literacy learning models made by recent science education researchers such as 21st-century learning (Suwono et al., 2017; Pujawan et al., 2022), contextual or culturally relevant pedagogy (Fakhriyah et al., 2017; Dewi et al., 2019; Hastuti et al., 2020), problem-posed instruction (Afriana et al., 2016; Parno et al., 2020), and technology-enhanced education (Ahied et al., 2020; Widodo et al., 2020).

In this study, our focus is not intended to characterise non-science majors differ from science majors. It can be a source of bias reported by a paper written by a particular science field. Thus, completing the current discussion with further study compared with science students must be highly recommended. Eventually, the random intercept model is determined by our study due to the limitation to harvesting higher levels of predictors provided by the university information system. Engaging another multilevel model with such a random slope would be strongly suggested.

Despite the potential issue reported above, a novel contribution made by this paper must be challenging. Theoretically, nested factors could be present within the educational setting and students' learning process. This study is evidence that science literacy educators should carefully emphasize the concept of authentic and heuristic assessment to provide science education more equitable for all students' aspects. Empirically, few demographic factors have significantly impacted scientific literacy education. The barriers driven by the construct of social-economic factors should not be a reason for Indonesian society to be desperate due to underperforming students in the PISA and TIMSS surveys. Optimistic vision must be encouraged to make a more scientifically literate society. Indeed, further investigation qualitatively, quantitatively, or combined to understand the possible reason that underlies the facts of underperformed Indonesian students in the past international survey (PISA and TIMSS) should be approached by the evaluation attempts considering either the nature of students' competence or the gap happened in the Indonesian evaluation system.

CONCLUSION

The Indonesian education system has made a progressive policy to boost the more scientifically literate community. Supporting this vision through the higher education sector has been implemented by this paper via a science literacy and technology course (MKU 6217) designed for non-science Indonesian undergraduate students. Students are generally nested within department and faculty settings in the schema of the college administrative system. Evaluating a science literacy course using multilevel models reveals some information regarding the effect of department and faculty settings on students' performance throughout the course. In this study, most of the demographic aspects are unable to significantly influence the mean difference in students' science literacy. Instead, formative assessment and affective measures are the most substantial factors that should be carefully considered in evaluating students' science literacy more equitable for the whole students. Evidence provided by this paper should be a recommendation to the higher institution in preparing effective science literacy education for the prospective teachers of the future Indonesian science education.

REFERENCES

Afriana, J., Permanasari, A., & Fitriani, A. (2016). Project-based learning integrated to stem to enhance elementary school students' scientific literacy. *Jurnal Pendidikan IPA Indonesia, 5*(2).

Ahied, M., Muharrami, L. K., Fikriyah, A., & Rosidi, I. (2020). Improving students' scientific literacy through distance learning with augmented reality-based multimedia amid the covid-19 pandemic. *Jurnal Pendidikan IPA Indonesia, 9*(4), 499-511.

Akram, S., & Pervaiz, Z. (2020). Mother's education as a predictor of individual's opportunities to learn and earn. *Elementary Education Online*, *19*(4), 879-884.

BAKK UNY. (2019). *Structure of Curriculum*. UNY. http://bakk.uny.ac.id/informasi/peraturan-akademik/struktur-kurikulum

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1).

Bellová, R., Balážová, M., & Tomčík, P. (2021). Are attitudes towards science and technology related to critical areas in science education? *Research in Science & Technological Education*, 1-16.

Briones, S. K. F., Dagamac, R. J. R., David, J. D., & Landerio, C. A. B. (2022). Factors Affecting the Students' Scholastic Performance: A Survey Study. *Indonesian Journal of Educational Research and Technology*, *2*(2), 97-102.

Cheema, J. R. (2017). Cross-country gender DIF in PISA science literacy items. *European Journal of Developmental Psychology, 16*(2), 152–166.

Creswell, J. C. (2015). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*. Sage Publications.

Dewi, C. A., Khery, Y., & Erna, M. (2019). An ethnoscience study in chemistry learning to develop scientific literacy. *Jurnal Pendidikan IPA Indonesia, 8*(2), 279-287.

Ding, L. (2019). Theoretical perspectives of quantitative physics education research. *Physical Review Physics Education Research, 15*(2).

Efthimiou, C. J., & Llewellyn, R. A. (2004). Cinema as a tool for science literacy. *ArXiv Preprint Physics/0404078*.

Ersan, O., & Rodriguez, M. C. (2020). Socioeconomic status and beyond a multilevel analysis of TIMSS mathematics achievement given student and school context in Turkey. *Large-Scale Assessments in Education, 8*(1), 1–32.

Fakhriyah, F., Masfuah, S., Roysa, M., Rusilowati, A., & Rahayu, E. S. (2017). Student's science literacy in the aspect of content science?. *Jurnal Pendidikan IPA Indonesia, 6*(1).

Finch, W. H., Bolin, J. E., & Kelley, K. (2016). *Multilevel Modeling Using R*. CRC Press.

Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M. (2014). Developing a Measure of Scientific Literacy for Middle School Students. *Science Education, 98*(4), 549-580.

Fortus, D., Lin, J., Neumann, K., & Sadler, T. D. (2022). The role of affect in science literacy for all. *International Journal of Science Education, 44*(4), 535–555.

Güre, Ö. B., Kayri, M., & Erdoğan, F. (2020). Analysis of Factors Affecting PISA 2015 Mathematics Literacy via Educational Data Mining. *Ted Eğitim Ve Bilim. 45*(1).

Hamper, B. C., & Meisel, J. W. (2020). Introducing Nonscience Majors to Science Literacy via a Laboratory and Lecture Beer Brewing Course. *Journal of Chemical Education, 97*(5), 1289–1294.

Hastuti, P. W., Setianingsih, W., & Anjarsari, P. (2020). How to develop students' scientific literacy through the integration of local wisdom in Yogyakarta on science learning? *Journal of Physics: Conference Series*, *1440*(1).

Hobson, A. (2003). Designing Science Literacy Course. In T. M. Cooney (Ed.), *Teaching Science in the Two-year College* (pp. 45–46). NSTA Press.

Hobson, A. (2008). The Surprising Effectiveness of College Scientific Literacy Courses. *The Physics Teacher, 46*(7), 404-406.

Impey, C. (2013). Science literacy of undergraduates in the United States. *Organizations, People and Strategies in Astronomy, 2*(2), 353–364.

Impey, C., Buxner, S., Antonellis, J., Johnson, E., & King, C. (2011). A twenty-year survey of science literacy among college undergraduates. *Journal of College Science Teaching, 40*(4), 31–37.

Kamba, A.H., Oladayo, A.A., & Fakai A.A. (2020). Factors and Effect of Poor Background on the Students' Academic Performance in Physics at Senior Secondary School in Birnin Kebbi Metropolis. *Saudi Journal of Engineering and Technology, 5*(3), 114-119.

Kanim, S., & Cid, X. C. (2020). Demographics of physics education research. *Physical Review Physics Education Research, 16*(2).

Luo, W., Li, H., Baek, E., Chen, S., Lam, K. H., & Semma, B. (2021). Reporting Practice in Multilevel Modeling: A Revisit After 10 Years. *Re-*

view of Educational Research, 91(3), 311-355.

Mahmudah, U., Lola, S. M., Fatimah, S., & Suryandari, K. C. (2022). Academic Resilience and Science Academic Emotion in Numeration under Online Learning: Predictive Capacity of an Artificial Neural Network. *Jurnal Pendidikan IPA Indonesia*, *11*(4).

Mahmudah, U., Chamdani, M., Tarmidzi, T., & Fatimah, S. (2020). Robust regression for estimating the impact of students' social behaviors on scientific literacy. *Cakrawala Pendidikan*, *39*(2), 293-304.

Marzulina, L., Pitaloka, N. L., Herizal, H., Holandyah, M., Erlina, D., & Lestari, I. T. (2018). Looking at the Link between Parents' Educational Backgrounds and Students' English Achievement. *Indonesian Research Journal in Education (IRJE), 2*(1), 51–60.

Masud, S., Mufarrih, S. H., Qureshi, N. Q., Khan, F., Khan, S., & Khan, M. N. (2019). Academic Performance in Adolescent Students: The Role of Parenting Styles and Socio-Demographic Factors – A Cross-Sectional Study from Peshawar, Pakistan. *Frontiers in Psychology, 10*.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language*, *94*, 305-315.

Miller, J. D. (1998). The measurement of civic scientific literacy. *Public understanding of science*, *7*(3), 203.

Mohammadpour, E., Shekarchizadeh, A., & Kalantarrashidi, S. A. (2015). Multilevel modeling of science achievement in the TIMSS participating countries. *The Journal of Educational Research*, *108*(6), 449-464.

Nurjati, N., Rahayu, E. Y., & Khabib, S. (2022). Technical Guidance for The Preparation Of AKM-Based National Assessment of English Learning for State Senior High School Teachers In Surabaya. *Kanigara, 2*(1), 216–224.

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education*, *105*(4), 653-680.

Parkinson, J., & Adendorff, R. (2004). The use of popular science articles in teaching scientific literacy. *English for Specific Purposes*, *23*(4), 379-396.

Parno, Yuliati, L., Hermanto, F. M., & Ali, M. (2020). A case study on comparison of high school students scientific literacy competencies domain in physics with different methods: PBL-stem education, PBL, and conventional learning. *Jurnal Pendidikan IPA Indonesia, 9*(2), 159-168.

Priyambodo, E. (2016). The effectiveness of collaborative academic online based learning through students' self-regulated learning. *Journal of Education and Learning (EduLearn)*, *10*(4), 405-410.

Pujawan, I.G.N., Rediani, N.N., Antara, I.G.W.S., Putri, N.N.C.A., & Bayu, G.W. (2022). Revised Bloom Taxonomy-Oriented Learning Activities to Develop Scientific Literacy and Creative Thinking Skills. *Jurnal Pendidikan IPA Indonesia. 11*(1), 47-60.

Queiruga-Dios, M. Á., López-Iñesta, E., Diez-Ojeda, M., Sáiz-Manzanares, M. C., & Dorrío, J. B. V. (2020). Citizen Science for Scientific Literacy and the Attainment of Sustainable Development Goals in Formal Education. *Sustainability, 12*(10), 4283.

Rachman, T. A., Latipah, E., Zaqiah, Q. Y., & Erihadiana, M. (2021). Curriculum Innovation to Improve Indonesian Education in PISA International Assessment in Disruptive Education Era. *ACM International Conference Proceeding Series*.

Ramdani, A., Jufri, A. W., Gunawan, G., Fahrurrozi, M., & Yustiqvar, M. (2021). Analysis of Students' Critical Thinking Skills in terms of Gender Using Science Teaching Materials Based on The 5E Learning Cycle Integrated with Local Wisdom. *Jurnal Pendidikan IPA Indonesia*, *10*(2), 187-199.

Ratini, M. H., Suparman, MA, Tamuri, AH, & Susanto, E. (2018). The influence of learning models and learning reliance. *Jurnal Pendidikan IPA Indonesia*, *7*(4), 458-466.

Rofieq, A., & Fauzi, A. (2022). Students' Knowledge and Attitudes toward Science: Its Correlation on Students' Disbelief in Non-Scientific Misinformation. *Jurnal Pendidikan IPA Indonesia*, *11*(2).

Romero, R., & De las Heras, M. A. (2020). One step closer to understanding motivation in scientific education: The incorporation of science textbooks as a predictor, together with motivational climate and basic psychological needs. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 590-599.

Ross, K., Hooten, M. A., & Cohen, G. (2013). Promoting Science Literacy through an Interdisciplinary Approach. *Bioscene: Journal of College Biology Teaching*, *39*(1), 21-26.

Rudolph, J. L. (2020). The lost moral purpose of science education. *Science Education*, *104*(5), 895-906.

Salamah, A., Prastowo, T., & Hariyono, E. (2022). The Development Research: Assessment Instruments of Science Literacy Based on Minimum Competency Assessment (AKM) Level 4th for Solar System Concepts. *Jurnal Penelitian & Pengembangan Pendidikan Fisika, 8*(1), 115–124.

Salehi, S., Burkholder, E., Lepage, G. P., Pollock, S., & Wieman, C. (2019). Demographic gaps or preparation gaps?: The large impact of incoming preparation on performance of students in introductory physics. *Physical Review Physics Education Research*, *15*(2), 020114.

Santoso, P. H., Istiyono, E., Haryanto, & Hidayatulloh, W. (2022). Thematic Analysis of Indonesian Physics Education Research Literature Using Machine Learning. *Data, 7*(11), 147.

Simmons, A. B., & Heckler, A. F. (2020). Grades, grade component weighting, and demographic disparities in introductory physics. *Physical Review Physics Education Research, 16*(2).

Sjöström, J., & Eilks, I. (2018). Reconsidering Different Visions of Scientific Literacy and Science Education Based on the Concept of Bildung. In Y. J. Dori, Z. R. Mevarech, & D. R. Baker (Eds.), *Cognition, Metacognition, and Culture in STEM Education: Learning, Teaching and Assessment* (pp. 65–88). Springer International Publishing.

Sumarti, S. S., Aris, S. R. S., & Aini, R. N. (2018). Chemoentrepreneurship with Cooperative Integrated Process Inquiry Strategy to Increase Students' Entrepreneurial Interest and Learning Motivation. *Jurnal Pendidikan IPA Indonesia*, *7*(2), 172-180.

Surjono, H. D., Muhtadi, A., & Wahyuningsih, D. (2017). The implementation of blended learning in multimedia courses for undergraduate students in Indonesia. International *Journal of Information and Education Technology, 7*(10), 783–786.

Surpless, B., Bushey, M., & Halx, M. (2018). Developing Scientific Literacy in Introductory Laboratory Courses: A Model for Course Design and Assessment. *Journal of Geoscience Education, 62*(2), 244–263.

Susongko, P., Arfiani, Y., & Kusuma, M. (2021). Determination of gender differential item functioning in tegal-students' scientific literacy skills with integrated science (Slisis) test using Rasch model. *Jurnal Pendidikan IPA Indonesia, 10*(2), 270-281.

Suwono, H., Pratiwi, H. E., Susanto, H., & Susilo, H. (2017). Enhancement of students' biological literacy and critical thinking of biology through socio-biological case-based learning. *Jurnal Pendidikan IPA Indonesia*, *6*(2), 213-220.

Ustun, U., Cansiz, M., Ozdemir, E., & Cansiz, N. (2022). Student and school-level factors to predict science literacy for two top-performing countries in PISA 2015: Finland and Singapore. *International Journal of Science Education, 44*(4), 579–603.

Widiana, I. W., Hermayuni, T. D., Sastra Agustika, G. N., & Kusumastuti, F. A. (2020). The effect of literacy based on the exploration of science with cultural insights on thematic content mastery and social attitude. *Jurnal Pendidikan IPA Indonesia, 9*(4), 521-531.

Widodo, W., Sudibyo, E., Suryanti, Sari, D. A. P., Inzanah, & Setiawan, B. (2020). The effectiveness of gadget-based interactive multimedia in improving generation z's scientific literacy. *Jurnal Pendidikan IPA Indonesia, 9*(2), 248-256.

You, H. S., Park, S., & Delgado, C. (2021). A closer look at US schools: What characteristics are associated with scientific literacy? A multivariate multilevel analysis using PISA 2015. *Science Education*, *105*(2), 406-437.