Assessing student understanding of measurement and uncertainty

S Jirungnimitsakul [1,2] and P Wattanakasiwich [1,2]

[1] Department of Physics and Materials Science, Chiang Mai University, Chiang Mai, 50000, Thailand
[2] Thailand Center of Excellence in Physics (ThEP), Commission on Higher Education, Bangkok 10400, Thailand

Abstract. The objectives of this study were to develop and assess student understanding of measurement and uncertainty. A test has been adapted and translated from the Laboratory Data Analysis Instrument (LDAI) test, consists of 25 questions focused on three topics including measures of central tendency, experimental errors and uncertainties, and fitting regression lines. The test was evaluated its content validity by three physics experts in teaching physics laboratory. In the pilot study, Thai LDAI was administered to 93 freshmen enrolled in a fundamental physics laboratory course. The final draft of the test was administered to three groups—45 freshmen taking fundamental physics laboratory, 16 sophomores taking intermediated physics laboratory and 21 juniors taking advanced physics laboratory at Chiang Mai University. As results, we found that the freshmen had difficulties in experimental errors and uncertainties. Most students had problems with fitting regression lines. These results will be used to improve teaching and learning physics laboratory for physics students in the department.

1. Introduction

Acquisition of data relate to the measurement is fundamental to science. Every measurement has an uncertainty associated with it; therefore, measured data have to be accurate and precise in order to provide correct interpretation. Most fundamental physics laboratory courses link theory to practice to develop experimental skill, involve measurement and interpretation of graph [1]. In a recent study, the freshmen who have completed fundamental physics laboratory courses were able to apply the routines for calculating means and standard deviations but were not able to interpret the spread in sets of repeated measurements and lacked an appreciation of uncertainty in experimental measurement [1-3].

Most fundamental physics laboratory courses in Thailand do not support students to design experiment in order to obtain accurate and reliable data. Students then often neglect the measurement uncertainty. Making a record of the data is just a number without considering the significant number. Therefore, in this paper, the objectives of this study were to develop and assess students understanding of measurement and uncertainty. The study focused on students majored in physics.

2. Methodology

2.1. Participants

The LDAI aims to assess students majoring in physics of their data analysis skills. The final version of Thai LDAI was administered to three groups of undergraduate majoring in physics at Chiang Mai University.

- Pilot group consisted of 93 freshmen taking a fundamental physics laboratory I.
- Group 1 consisted of 45 freshmen. They were currently taking a fundamental physics laboratory II.
- Group 2 consisted of 16 sophomores. They already took intermediate physics laboratory I and were currently taking intermediated physics laboratory II.
- Group 3 consisted of 21 juniors. They were currently taking advanced physics laboratory.

Participants in three groups were volunteers with compensation for participating in the studies.

2.2. Conceptual test

The instrument has been translated and adapted from Laboratory Data Analysis Instrument (LDAI) test [4], consisting of 30 items (19 multiple-choice and 11 true/false questions). The LDAI focused on four objectives including (1) measures of central tendency, (2) experimental errors and uncertainties, (3) fitting regression lines and (4) graph and aims. In the original LDAI, the experimental context was a cart traveling on a smooth track by pulling force of forces. The cart's acceleration was measured by using a sensor. Main results were different values of measured acceleration when changing the magnitude of pulling forces. In the Thai LDAI, the experimental context was a simple harmonic motion of a mass hanging from a spring. This experiment is commonly found in a fundamental physics laboratory in Thailand.

The Thai LDAI instrument was evaluated its content validity by three physics experts. The first version of Thai LDAI (LDAI 1.0) consisted of 14 multiple-choice questions and 6 true/false questions, excluding questions on the objective 4. The test was administered to the pilot group. After performing item analysis, 10 questions with low difficulty index and low discrimination index were revised. The final-version of Thai LDAI (LDAI 2.0) was included questions on the objective 4, so it consisted of 15 multiple-choice questions and 10 true/false questions.

3. Results and discussion

3.1. Item and overall analysis

Only group 1 responses were used in determine item analysis and overall analysis. This is because the group 1 had the highest number of participants**.** Item analysis composed of three measures—item difficulty index (P-index), item discrimination index (D-index) and point biserial coefficient (PBI). Overall analysis of the survey consisted of Kuder-Richardson reliability (KR-20) and Ferguson's delta (δ) [5].

From Table 1, the item analysis including average P-index, D-index, and PBI were higher than the desired values. The LDAI 2.0 is considered to be a medium difficulty test and has satisfactory discrimination index and reliability for individual items. For the overall analysis, the instrument has satisfactory discrimination ability but the quite low reliability of the whole test as indicated by a KR-20 value.

Although the average P-index, D-index and PBI were higher than desired values, there were several questions having low values of P-index and PBI (< 3.0) as shown in figure 1. The questions with both low P-index and PBI will be discussed in more details on section 3.3.

Table 1. Item analysis and reliability of Thai LDAI 2.0

Analysis	LDAI 2.0 values	Desired values [5]
P-index	0.55	0.30≤P≤0.90
D-index	0.32	≥ 0.30
PBI	0.29	≥ 0.20
KR-20	0.59	≥ 0.70
δ	1.04	≥ 0.90

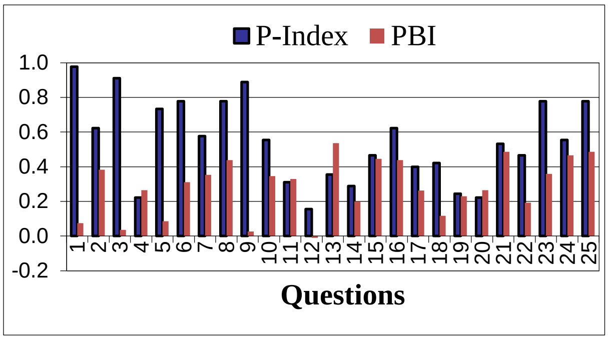

Figure 1. The P-index and PBI for each question

3.2. Descriptive statistical analysis

Overall, all three groups scored an average of correct answer higher than 50%. Sophomores had highest average of 68.00% correct answers. The average scores (M in %) and standard deviation (SD) are presented in Table 2 by objectives. All groups performed lowest scores on objective 3, fitting regression lines. Students might just understand regressions in theory, but they did have difficulties in applying it to the real context [1].

Table 2. Percentages of the correct student responses according to four objectives

Objectives	Group 1 (*N = 45*)		Group 2 (*N = 16*)		Group 3 (*N = 21*)	
	M	SD	M	SD	M	SD
(1) Measures of central tendency (1,2,3,10.1,11.1)	64.44	17.52	71.25	16.28	62.86	15.86
(2) Experimental errors and uncertainties (4-9,10.2,11.2,12,13)	54.00	17.24	73.75	19.96	68.57	17.11
(3) Fitting regression lines (14,15,16.1,16.2,16.3)	38.22	22.08	57.50	21.76	47.62	24.88
(4) Graphs and aims (17, 18.1-18.4)	62.22	27.38	63.75	22.17	71.43	24.96
Overall test	**54.58**	**13.38**	**68.00**	**10.01**	**63.81**	**11.76**

3.3. Student difficulties

Percentages of correct responses in each item according to each group are listed in Table 3. To better understand student difficulties, questions with low percentages (<40%) will be further analyzed. These include questions 11.1, 16.2 and 16.3. Also, freshmen had low scores on questions 4, 10.2 and 12. These are about experimental errors and uncertainties. Results of student difficulties inferred from dominated incorrect choices and their written explanations in each question.

- *Difficulties with fitting regression lines*

In question 16.2, most students chose to fit non-linear regression line for a force versus displacement graph. This indicated their lack of basic understanding on spring force. In question 16.3, most students did not think about error bar of data point and the regression line. The line should pass within one error bar of the data point.

- *Difficulties with experimental errors and uncertainties*

Freshmen had difficulties on this topic because they haven't had an intermediate physics laboratory, offered in the second year of undergraduate curriculum. This course emphasized on measurements and uncertainty, as well as, error analysis.

Most freshmen were confused between systemic error and random error. In question 11.2, freshmen incorrectly answered that the statistical error will remain unchanged although the numbers of repeated measurement increase. Also, freshmen did not understand distribution of data and average value. In question 11.1, most students did not understand that the average will remain unchanged when a 6th measurement has the same value as the average of the previous 5 measurements. These findings of student difficulties are similar to previous studies [1, 6]

Table 3. The percentage of correct answers in each item according to each group

Items	% correct answers			Items	% correct answers		
	Freshmen	Sophomores	Juniors		Freshmen	Sophomores	Juniors
1	97.78	100.00	95.24	12*	28.89	62.25	61.90
2	62.22	68.75	61.90	13	46.67	81.25	76.19
3	91.11	93.75	71.43	14	62.22	93.75	71.43
4*	22.22	56.25	61.90	15	40.00	62.50	61.90
5	73.33	62.50	76.19	16.1	42.22	62.50	28.57
6	77.78	93.75	85.71	16.2*	24.44	37.50	28.57
7	57.78	87.50	71.43	16.3*	22.22	31.25	47.62
8	77.78	93.75	80.95	17	53.33	75.00	66.67
9	88.89	93.75	80.95	18.1	46.67	43.75	52.38
10.1	55.56	75.00	76.19	18.2	77.78	81.25	95.24
10.2*	31.11	56.25	61.90	18.3	55.56	50.00	61.90
11.1*	15.56	18.75	9.52	18.4	77.78	68.75	80.95
11.2*	35.56	50.00	28.57	**Average**	**54.58**	**68.00**	**63.81**

3.4. Comparison among three groups
Levene test showed that each group had a normal distribution of LDAI scores. Post hoc multiple comparison using Fisher's least significant difference technique indicated significant differences between the means of freshmen versus sophomores ($p = 0.001$) and freshmen versus juniors ($p = 0.006$), but no significant difference between the means of sophomores and juniors ($p = 0.312$).

4. Conclusions
The Thai LDAI 2.0 was developed and implemented to assess physics students' concepts of data analysis. Most undergraduates had difficulties in the fitting regression line. Freshmen also had difficulties in experiment errors and uncertainties. From performing multiple comparisons among three groups, the freshmen had significantly lower average scores than the sophomores and the juniors. This might be because they haven't had the in-depth learning experience about data analysis in the introductory physics laboratory. Generally, this laboratory is taught using a "cookbook" style; as a result, students did not gain essential data analysis skills. Also, sophomores and juniors had difficulties with regression lines, so the intermediate laboratory should include more practices on this topic.

Refferences

[1] Kung R L 2005 *Am. J. Phys.* **73(8)** 771
[2] Allie S, Buffler A, Campbell B, Lubben F, Evangelinos D, Psillos D and Valassiades O 2003 *Phys. Teach.* **41** 23
[3] Volkwyn T S, Allie S and Buffler A 2008 *Phys. Edu. Res.* **4** 010108
[4] Eshach H and Kukliansky I 2016 *Can. J. Phys.* **94(11)** 1205
[5] Ding L and Beichner R. 2009 *Phys. Rev. ST Phys. Educ. Res.* **5** 020103
[6] Allie S, Buffler A, Kaunda L, Campbell B and Lubben F 1998 *Int. J. Sci. Educ.* **20(4)** 447