Kasetsart Journal of Social Sciences 46 (2025) 460124

Kasetsart Journal of Social Sciences

journal homepage: http://kjss.kasetsart.org

Development of Work and Energy in Multi-tiers Test (WE-MusT) to identify students' conceptions levels: A Rasch analysis model

Achmad Samsudin[a,*], Dinda Lestari[a], Ida Kaniawati[a], Andi Suhandi[a], Hadi Nasbey[b], Nurjannah Nurjannah[c], Adam Hadiana Aminudin[d], Asep Dedy Sutrisno[e], Firmanul Catur Wibowo[f], Siswanto[g], Binar Kurnia Prahani[h], Bayram Coştu[i]

[a] *Department of Physics Education, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, West Java 40154, Indonesia*
[b] *Department of Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, East Jakarta, Special Capital Region of Jakarta 13220, Indonesia*
[c] *Physics Education, Faculty of Teacher Training and Education, Universitas Tadulako, Palu, Central Sulawesi 94148, Indonesia*
[d] *Physics, Faculty of Mathematics and Sciences, Universitas Kebangsaan Republik Indonesia, Bandung, West Java 40263, Indonesia*
[e] *Public High School 1 Karangkobar, Karangkobar, Banjarnegara, Central Java 53453, Indonesia*
[f] *Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta, East Jakarta, Jakarta 13220, Indonesia*
[g] *Science Education, Faculty of Teacher Training and Education, Universitas Tidar, Magelang, Central Java 56116, Indonesia*
[h] *Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya, Surabaya, East Java 60231, Indonesia*
[i] *Department of Science Education, Faculty of Education, Yildiz Technical University, Yıldız, 34349 Beşiktaş/İstanbul, Turkey*

Article Info

Article history:
Received 6 September 2022
Revised 16 May 2024
Accepted 20 May 2024
Available online 28 March 2025

Keywords:
conception,
Multi-tiers test,
Rasch analysis,
WE-MusT,
work and energy

Abstract

This study aims to develop the Work and Energy in Multi-tiers Test (WE-MusT) to Identify students' conceptions levels. The research method has been implemented in 4D (Define, Design, Develop, Disseminate). The participants in this study were 230 students, namely, 81 male participants and 149 female participants with an average age of 15-18 years in several high schools and their equivalents in Java. Meanwhile, the research instrument is the product developed in this study. Rasch model analysis was used to identify the level of conceptions. Based on data analysis, all items can be used, except S1 and S6 (Invalid). The WE-MusT also has reliability (0.75) with a Cronbach Alpha, meanwhile, conceptions level of Sound Understanding (14%), Misconception (37%), Partial Negative (31%), Partial Positive (1%), No Understanding (9%), and No Coding (8%). Based on these results, it can be concluded that many students still experience misconceptions and do not understand the concepts of work and energy. This research has implications for the quality of instruments in identifying students' conception levels and can be a reference in developing diagnostic test instruments.

* Corresponding author.
E-mail address: achmadsamsudin@upi.edu (Samsudin, A.).

https://doi.org/10.34044/j.kjss.2025.46.1.24
2452–3151/© 2025 Kasetsart University.

Introduction

Multi-tiers test is a multiple-choice diagnostic test instrument with a level of confidence. The multiple choice developed is that there are four levels. Levels one and three are multiple choice, while levels two and four are confidence levels (Aminudin et al., 2019; Samsudin et al., 2025; Yang et al., 2018). This test instrument is used to measure student knowledge, understanding, and confidence level. Using these test instruments also can identify the level of scientific knowledge, lack of knowledge and misconceptions (Kiray & Simsek, 2021; Osman et al., 2017; Yen et al., 2004). The development of this diagnostic instrument is important, because misconceptions must be identified and addressed immediately. This misconception can occur in any concept, one of which is in physics.

Physics is a branch of science that studies objects, phenomena and interactions between objects in nature. The matter of work and energy is one of the concepts of physics that is very important to study because it is related and found in everyday life (Chabalengula et al., 2012). The understanding of the concept of work and energy is still low, and some students experience misconceptions. As many as 73 percent of students experience misconceptions about work and energy (Samsudin et al., 2021). Students can solve work and energy problems using mathematical formulas, but it is still difficult to understand concepts. As many as 65 percent of students have a sufficient understanding of concepts, while 35 percent of students do not understand concepts in work and energy concepts (Zulfa et al., 2019). However, the concept of work and energy in physics is still broad and has the potential to develop diagnostic tests. Thus, we conducted this study to develop the Work and Energy in Multi-tiers Test (WE-MusT) to Identify students' conceptions levels using Rasch analysis.

Rasch analysis is a probabilistic model developed by Danish mathematician Georg Rasch in 1960 (Boone et al., 2011; Wei et al., 2014; Yang et al., 2018). This Rasch analysis describes the relationship between individuals and items. Rasch analysis is used to analyze tests, items, and misconceptions in physics learning (Eliya et al., 2024; Ferretti & Giberti, 2021; Samsudin, Azizah, et al., 2024). Rasch analysis can also be used to develop test items, providing relevant information related to the assessment of students during the learning process (Samsudin, Rusdiana, et al., 2021). Using Rasch analysis in this study can maximize the quality of the instruments developed.

Methodology

Research Design

The research design used in this study is 4D, namely, Define, Design, Develop, and Disseminate. The 4D design was chosen because it was in accordance with the purpose of this research, namely, development research. At the Define stage is a literature study related to the diagnostic test instrument and the level of conception of students. In the Design stage, researchers prepare question instruments which include making a distribution of question grids, question indicators, and making question items. In the Develop stage, researchers validate the expert of the instrument carried out by the validator to find out the roughness of the question item, then analyze the validation results and revise the instrument according to the direction of the validator. At the Disseminate stage, we give WE-MusT to students.

Participants

This research was conducted in several high schools in Java. The study sample amounted to 230 students of class XI-XII with an average of 15-18 years. The total number of participants is 81 males and 149 females. Table 1 shows the distribution of participants for each province.

Table 1 Distribution of participants in each province

Province (city/districts)	Number of Participant	School of Distribution
Banten (1. Serang, 2. Tangerang)	37	2
Jakarta (3. West jakarta, 4. South jakarta)	16	2
West Java (5. Sukabumi, 6. Kuningan)	77	3
Central Java (7. Banjarnegara)	47	1
Yogyakarta (8. Sleman)	2	1
East Java (9. Kediri, 10. Tuban)	51	3

Based on Table 1, the number of distributions for each province is disproportionate because there are instruments spread by researchers through several intermediaries, causing communication constraints between researchers and participants. However, this can be overcome with good communication when distributing the instrument, thus, it does not affect the quality of the instrument being developed. Following is a map of the location of the development of the work and energy diagnostic test instrument in Figure 1.

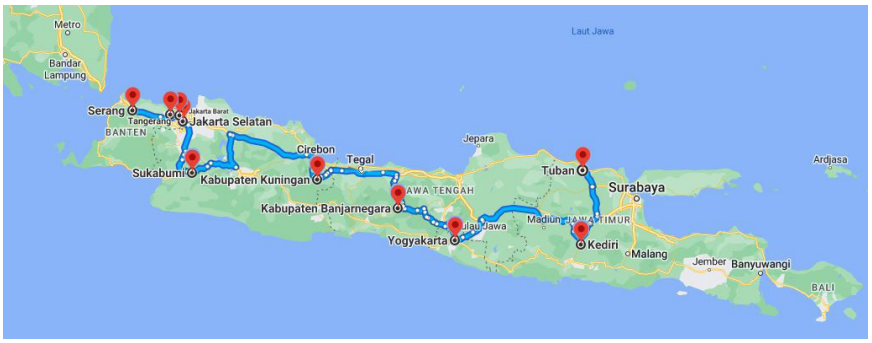

Figure 1 Map location of the instrument development (Source by Google maps)

Data Analysis

Data analysis was carried out using Rasch analysis assisted by Winstep 4.4.5 software with coding techniques. The level of conception in students consists of Sound Understanding (SU), Partial Positive (PP), Partial Negative (PN), No Understanding (NU), Misconception (MC), and No Coding (NC) (Aminudin et al., 2019). Table 2 shows the categories of students' conception levels.

The scores in Table 2 are not like the conception category scores in general, and seem to be reversed. This was indeed done because the score for the MC category must be the largest, thus, when analyzed using Rasch analysis, it can be identified, likewise, for other scores such as NU, PN, PP, and SU, which successively become smaller scores. After categorizing and scoring, the next step is Rasch analysis using the output for Item Fit Order, Reliability, and Wright Maps. The use of Wright Maps can help in mapping students' abilities towards the WE-MusT items being developed. Moreover, Rasch analysis uses logarithms in its scaling and has also been used by several researchers regarding mapping students' conceptions (Aminudin et al., 2019; Amiruddin et al., 2023; Kurniawan et al., 2023; Samsudin et al., 2023). Thus, the Rasch analysis can complement our views in assessing students' conceptions.

Table 2 Conception level categories

Category	Score	No	Tier 1	Tier 2	Tier 3	Tier 4
SU	0	1.	T	S	T	S
PP	0	2.	T	U	T	S
		3.	T	S	T	U
		4.	T	U	T	U
PN	1	5.	T	S	F	S
		6.	T	U	F	S
		7.	T	S	F	U
		8.	T	U	F	U
		9.	F	S	T	S
		10.	F	U	T	S
		11.	F	S	T	U
		12.	F	U	T	U
NU	3	13.	F	U	F	S
		14.	F	S	F	U
		15.	F	U	F	U
MC	4	16.	F	S	F	S
NC	(empty)	17.	NCA			

Note: T = True; F = False, S = Sure; U = Unsure; NCA = No Clear Answer

Results and Discussion

The results in this study will be discussed based on a 4D (Define, Design, Develop, Disseminate) design. This is done for the explanation can be more systematic. The explanation is discussed based on the sub of the design.

Define Stage

In the define stage, is the initial identification of students' misconceptions with determining the indicators of work and energy misconceptions and literature study. Identification is carried out for sub-concepts that still have misconceptions in them. Some misconceptions are presented in Table 3. After identification is done, the results will be compiled for the design stage.

Design Stage

The design phase begins with making a grid distribution of the question instruments, compiling question indicators, and making questions. The creation of indicators of misconceptions cannot be separated from the literature study that has been carried out. The distribution of the question instrument grid is made so that the items are distributed evenly. The following is Table 4 Instrument grid distribution format.

Table 3 Misconception of work and energy

Sub Concept	No	Misconception
Work	1.	Work is positive if it is to the right, while work is negative if it is to the left
	2.	Work can't be determined directly through graphs.
	3.	The least work is obtained if the trajectory traversed by the moving object is small
	4.	In a non-conservative force, work is independent of the trajectory
Kinetic Energy	5.	The kinetic energy of the object when it pounds the ground is zero.
	6.	Kinetic energy is affected by the slope of a field or the trajectory of the object
Velocity of object on slope field	7.	The velocity of the object at every point on the slope field is the same.
Potential Energy	8.	Potential energy will be smaller if the trajectory is getting shorter
	9.	Potential energy will be greater if the trajectory is getting longer or more difficult
Mechanical Energy	10.	An object can traverse a trajectory depending on the steepness of the trajectory
	11.	When there is a change in kinetic energy, the mechanical energy will also change.

Table 4 Format of the grid distribution

Sub Concept	Number of Misconception	Item Question	Question Construction	
			Descriptor	Option
Work	1.	S1	F-S	F
	2.	S2	S-F	MS
	3.	S3	S-F	S
	4.	S10	S-F-S	MS
Kinetic Energy	5.	S4	F-S	F
	6.	S5	S-F	MS
	7.	S6	S-F	MS
Potential Energy	8.	S7	F-S	S
	9.	S11	S-F	S
Mechanical Energy	10.	S8	F	S
	11.	S9	S-F	D

Note: S = Statements, F = Figures, T = Tables, D = Diagrams, MS = Mathematical Symbols

Develop Stage

At the develop stage, the instrument preparation starts from a 3 Tier instrument to a 4 Tier instrument (Work and Energy in Multi-tiers Test (WE-MusT)). Tier 1 is multiple choice, Tier 2 is belief in Tier 1 choice answers, Tier 3 is the reasons for the answers that have been chosen in Tier 1, and Tier 4 is the level of confidence in Tier 3 reasons. Figure 2 below is the development of multi-tier type 4 Tier.

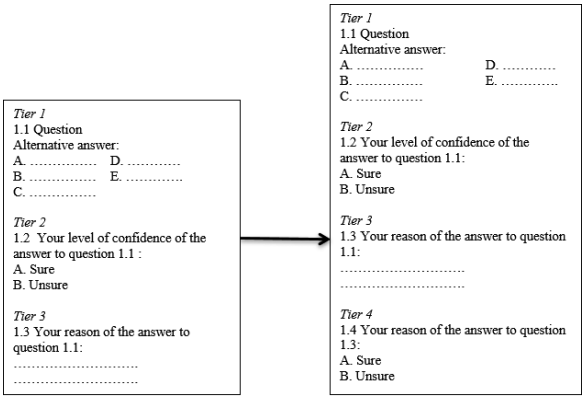

Figure 2 (a) Three tier tests, (b) WE-MusT

Disseminate Stage

The dissemination stage was carried out in several areas on the island of Java (Table 1). The WE-MusT are distributed via google form. Once sufficient, then the data from the distribution of WE-MusT are processed using the Rasch analysis model, such as analysis of the instrument (Item Analysis and Reliability) and analysis in identifying student conceptions (Wright Maps and Students' Conception Level).

Item analysis

The fit item is used to identify the functionality of the question item in the measurement. Items are said to be appropriate if the value of Outfit and Infit MNSQ is $0.5 < MNSQ < 1.5$; if the value obtained is more than 1.5, then the item is potentially incompatible with the Rasch model and is considered for deletion (Prieto et al., 2003; Smith et al., 2008). The Z-Standard Outfit value (ZSTD) is $-2 < ZSTD < +2$, and the value of Point Measure Correlation (Pt. Mean Corr) is $0.32 <$ Pt. Mean Corr < 0.85. Items are said to be fit if they meet a minimum of one of the three categories (Kurnianto et al., 2023).

Figure 3 shows the results of Rasch analysis processing on fit order items. There are MNSQ, ZSTD and Pt Mean Corr Outfit values in the fit order item processing results. Table 5 presents the results of processing fit order items to determine the item's suitability.

Based on Table 4, items S1 and S6 are invalid and both are not used for further analysis calculations.

Reliability

The analysis of the instrument for the reliability value is shown in Figure 4.

Figure 4 shows a Cronbach Alpha value of 0.75, indicating good reliability in the data. Cronbach Alpha values are used to measure interactions between individuals and items. Cronbach Alpha values are categorized as good if they are 0.7–0.8 (Hansson et al., 2021; Julianto et al., 2025; Wati et al., 2019). Meanwhile, grouping Person and Item is done by calculating the SEPARAION value for Equation (1) and the results can be seen in Table 6.

$$H = \frac{[4xSEPARATION]+1}{3} \quad (1)$$

INPUT: 230 PERSON 11 ITEM REPORTED: 230 PERSON 11 ITEM 5 CATS WINSTEPS 4.4.5

PERSON: REAL SEP.: 1.19 REL.: .59 ... ITEM: REAL SEP.: 6.07 REL.: .97

ITEM STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PTMEASUR-AL CORR.	PTMEASUR-AL EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	ITEM
6	550	207	-.27	.05	1.23	3.01	1.33	2.44	A .34	.45	22.0	22.1	S6
4	522	216	-.15	.05	1.16	2.20	1.23	1.96	B .35	.49	21.9	22.4	S4
3	614	207	-.44	.05	1.16	1.80	1.19	1.20	C .38	.43	24.9	30.1	S3
11	445	208	.01	.05	1.18	2.37	1.14	1.38	D .48	.49	14.1	19.1	S11
10	452	207	-.01	.05	.94	-.85	1.09	.85	E .51	.49	17.1	19.8	S10
9	646	204	-.56	.06	.81	-2.01	1.07	.43	F .37	.40	45.0	38.8	S9
2	388	225	.23	.05	1.02	.28	1.04	.40	e .54	.50	26.0	31.3	S2
7	335	215	.33	.05	1.01	.13	.97	-.26	d .59	.51	23.5	33.8	S7
8	435	211	.06	.05	.84	-2.21	.99	-.03	c .51	.49	23.4	22.4	S8
5	414	212	.11	.05	.80	-2.80	.83	-1.75	b .51	.50	36.2	28.0	S5
1	239	226	.69	.06	.78	-1.99	.63	-2.76	a .64	.49	47.3	43.6	S1
MEAN	458.2	212.5	.00	.05	.99	.0	1.05	.4		27.4	28.3		
P.SD	114.0	7.0	.34	.00	.16	2.0	.18	1.5		10.3	7.7		

Figure 3 Item Fit Order

Based on Table 6, it can be identified that people can be grouped into 2 categories, and items can be grouped into 7 categories.

Wright maps

Wright Maps determine the relationship between item difficulty and student ability (Salimpour et al., 2022). The left side of the wright maps is the level of difficulty of items, while the right side is the level of ability of students to answer items. Students at the top indicate that the learner has a very high potential for misconceptions, while those at the bottom show that the learner has a very low potential for misconceptions. The top item has good quality in measuring misconceptions, while the bottom item is the item that is slightly answered by the learner (Aminudin et al., 2019).

SUMMARY OF 230 MEASURED (EXTREME AND NON-EXTREME) PERSON

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	18.5	8.3	.29	.32				
SEM	.5	.1	.04	.02				
P.SD	8.0	1.8	.55	.23				
S.SD	8.0	1.8	.55	.23				
MAX.	36.0	9.0	2.62	2.08				
MIN.	.0	1.0	-1.57	.23				

REAL RMSE .41 TRUE SD .36 SEPARATION .89 PERSON RELIABILITY .44
MODEL RMSE .39 TRUE SD .38 SEPARATION .97 PERSON RELIABILITY .49
S.E. OF PERSON MEAN = .04

PERSON RAW SCORE-TO-MEASURE CORRELATION = .81
CRONBACH ALPHA (KR-20) PERSON RAW SCORE "TEST" RELIABILITY = .75 SEM = 4.02

SUMMARY OF 9 MEASURED (NON-EXTREME) ITEM

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	472.3	211.7	.00	.05	1.00	-.09	1.05	.35
SEM	34.3	2.1	.10	.00	.06	.70	.05	.47
P.SD	97.0	6.0	.28	.00	.16	1.99	.15	1.33
S.SD	102.9	6.4	.30	.00	.17	2.11	.16	1.41
MAX.	646.0	225.0	.38	.06	1.23	2.93	1.36	3.06
MIN.	335.0	204.0	-.52	.05	.81	-2.52	.83	-1.79

REAL RMSE .05 TRUE SD .28 SEPARATION 5.10 ITEM RELIABILITY .96
MODEL RMSE .05 TRUE SD .28 SEPARATION 5.28 ITEM RELIABILITY .97
S.E. OF ITEM MEAN = .10

ITEM RAW SCORE-TO-MEASURE CORRELATION = -1.00
Global statistics: please see Table 44.
UMEAN=.0000 USCALE=1.0000

Figure 4 Summary Statistic

Table 5 Format of the grid distribution

No. Question	Score of MNSQ	Score of ZSTD	Score of Pt Mean Corr	Availability Status (Category/s)	Interpretation
S1	0.63	-2.76	0.64	2	*Invalid*
S2	1.04	0.40	0.54	3	Valid
S3	1.19	1.20	0.38	3	Valid
S4	1.23	1.96	0.35	3	Valid
S5	0.83	-1.75	0.51	3	Valid
S6	1.33	2.44	0.34	2	*Invalid*
S7	0.97	-0.26	0.59	3	Valid
S8	0.99	-0.03	0.51	3	Valid
S9	1.07	0.43	0.37	3	Valid
S10	1.09	0.85	0.51	3	Valid
S11	1.14	1.38	0.48	3	Valid

Table 6 SEPARATION Calculation Results

Criteria				Data	Results
EXTREME AND	Person	Real	SEPARATION	0.89	2
NON-EXTREME		Model	SEPARATION	0.97	2
NON-EXTREME	Item	Real	SEPARATION	5.1	7
		Model	SEPARATION	5.28	7

Figure 5 shows that students in numbers 004W (West Java), 139E, 170E (East Java) and 216B (Banten) have the potential to answer with the most misconceptions for all items because these two numbers are at the top of the Wright maps. Meanwhile, the number 064W (West Java) has the potential to answer with the fewest misconceptions because the number is at the bottom of the Wright maps. Item number S7 is at the very top, which indicates that the item has the most potential to be answered with misconceptions by students, while item number S9 is at the bottom, which indicates that the item has the potential to be answered with the least misconceptions by students. Overall, the item has potential answered misconceptions to students.

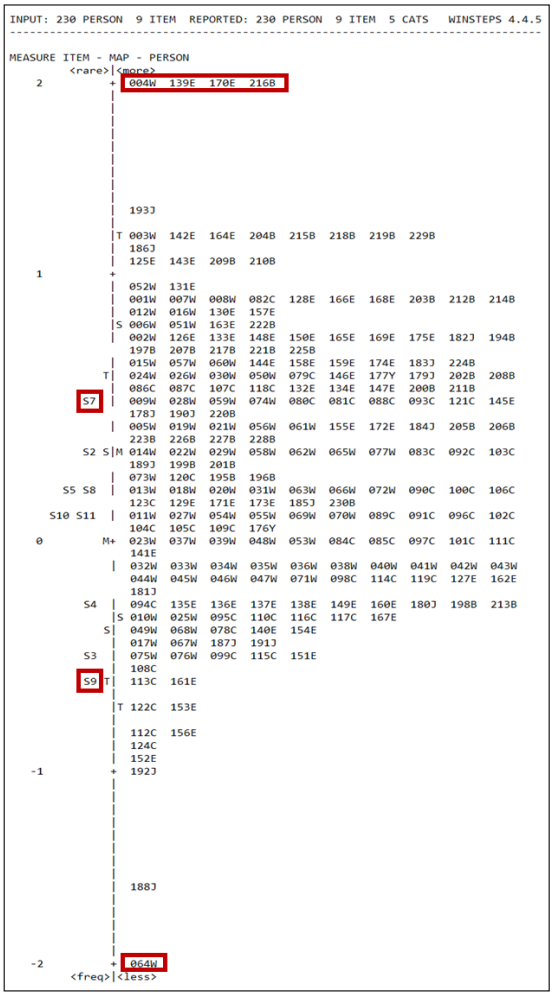

Figure 5 Variable (Wright) Maps

Students' conception level

Based on the conception levels, understanding students' conceptions are different. The concept becomes one of the most important aspects for students (Amiruddin et al., 2025; Samsudin et al., 2025; 2024). Table 7 shows the results of the recapitulation of the level of conception in students.

Based on Table 7, the highest percentage in the MC category is in the S3 questions (57%) and the lowest in the S2 questions (26%). In the NU category, the highest percentage is on S9 questions (23%) and the lowest is on S7 questions (3%). For the PN category, the highest percentage is on S5 questions (54%) and the lowest is on S9 questions (14%). In the PP category, the highest percentage is in S7 questions (2%) and the lowest is in S3 and S9 questions (0%). For the SU category, the highest percentage was in S7 questions (30%) and the lowest was in S9 questions (2%). Finally, in the NC category, the highest percentage was on S9 questions (11%) and the lowest was on S2 questions (2%). Meanwhile, the mean for each concept is shown by the distribution diagram of the students' level of conception in Figure 6.

The results in the Figure 6 show that the students' abilities are the highest in the MC category (37%) and the lowest in the PP category (1%). This indicates that students still experience many misconceptions about the concept of work and energy. The potential causes of misconceptions include the condition of students, teachers, teaching methods, books, low students' absorption, complex and abstract concept. Research by Osman et al. (2017), Kaniawati et al. (2019), Fratiwi et al. (2020) declare that several factors' misconception are low absorption of students and abstract concept, and this makes it all make sense why there are still many misconceptions about the concepts of work and energy. For this reason, the results of this research show that future research is needed to think about ways to address these misconceptions. Several methods such as developing and implementing learning models, developing media (hands on or virtual), and several learning tools can be alternative solutions to this problem.

Table 7 The recapitulation of conception levels

No	Sub Concept	Item	Frequency and percentage of each level of students											
			MC		NU		PN		PP		SU		NC	
			f	%	*f*	%	*f*	%	*f*	%	*f*	%	*f*	%
1	Work	S2	60	26	26	11	70	30	2	1	67	29	5	2
		S3	132	57	17	7	35	15	1	0	22	10	23	10
		S10	78	34	23	10	71	31	2	1	33	14	23	10
2	Kinetic energy	S4	96	42	17	7	87	38	2	1	15	7	14	6
		S5	65	28	10	4	124	54	4	2	9	4	18	8
3	Potential energy	S7	61	27	6	3	73	32	5	2	70	30	15	7
		S11	89	39	12	5	53	23	3	1	50	22	22	10
4	Mechanical	S8	71	31	17	7	100	43	2	1	21	9	19	8
	Energy	S9	113	49	54	23	32	14	0	0	5	2	26	11

Notes: S = Question, MC = Misconception, NU = No Understanding, PN = Partial Negative, PP = Partial Positive, SU = Sound Understanding, NC = No Coding

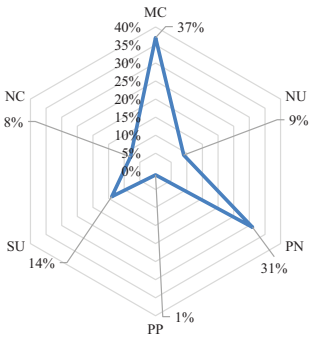

Figure 6 Distribution of conception levels

Conclusion

The results of Rasch analysis for instrument development show that all items can be used. The WE-MusT also have reliability with a Cronbach Alpha value of 0.75 (good). Meanwhile, in identifying students' conceptions, WE-MusT can identify the conceptions level of Sound Understanding with a percentage of 14 percent, Misconception with a percentage of 37 percent, Partial Negative with a percentage of 31 percent, Partial Positive with a percentage of 1 percent, No Understanding with a percentage of 9 percent, and No Coding with a percentage of 8 percent. Based on the results, the most significant percentage in the MC category of 37 percent; it can be concluded that many students still experience misconceptions and do not understand the concepts of work and energy. Thus, WE-MusT is feasible to be used to identify students' level of conception. These results can be a tool that can be used by practitioners in the field. In addition, this research has implications for the quality of instruments in identifying students' conception levels and can be a reference in developing diagnostic test instruments.

Conflict of interest

The authors declare that there is no conflict of interest.

Acknowledgments

Thanks to RKLI Litabmas Universitas Pendidikan Indonesia (UPI) - Universitas Negeri Jakarta (UNJ) and DPPM Kemdiktisaintek which has supported the funding of this research.

References

Aminudin, A. H., Kaniawati, I., Suhendi, E., Samsudin, A., Coştu, B., & Adimayuda, R. (2019). Rasch analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): developing and assessing second-years Sundanese-scholars alternative conceptions. *Journal for the Education of Gifted Young Scientists*, *7*(3), 607–629. https://doi.org/10.17478/jegys.574524

Amiruddin, M. Z. Bin, Samsudin, A., Suhandi, A., Coştu, B., & Prahani, B. K. (2025). Scientific mapping and trend of conceptual change: A bibliometric analysis. *Social Sciences and Humanities Open*, *11*(November 2024). https://doi.org/10.1016/j.ssaho.2024.101208

Amiruddin, M. Z. Bin, Samsudin, A., Suhandi, A., Kaniawati, I., Coştu, B., Aminuddin, A. H., & Kuniawan, F. (2023). Validity and reliability of the global warming instrument: A pilot study using Rasch model analysis. *Jurnal Pendidikan MIPA*, *24*(4), 911–922. http://dx.doi.org/10.23960/jpmipa/v24i4.pp912-922

Boone, W. J., Townsend, J. S., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, *95*(2), 258–280. https://doi.org/10.1002/sce.20413

Chabalengula, V. M., Sanders, M., & Mumba, F. (2012). Diagnosing students' understanding of energy and its related concepts in biological context. *International Journal of Science and Mathematics Education*, *10*(2), 241–266. https://doi.org/10.1007/s10763-011-9291-2

Eliya, M., Purwianingsih, W., Supriatna, A., Samsudin, A., & Hadiana, A. (2024). Electromotive Forces on Multitier Instruments (EFoMI): Development to measure student misconception with Rasch. *Journal of Environment and Sustainability Education*, *2*(1), 40–52. https://doi.org/10.62672/joease.v2i1.32

Ferretti, F., & Giberti, C. (2021). The Properties of Powers: Didactic Contract and Gender Gap. *International Journal of Science and Mathematics Education*, *19*(8), 1717–1735. https://doi.org/10.1007/s10763-020-10130-5

Fratiwi, N. J., Samsudin, A., Ramalis, T. R., Saregar, A., Diani, R., Irwandani, Rasmitadila, & Ravanis, K. (2020). Developing memori on Newton's laws: For identifying students' mental models. *European Journal of Educational Research*, *9*(2), 699–708. https://doi.org/10.12973/eu-jer.9.2.699

Hansson, L., Hansson, Ö., Juter, K., & Redfors, A. (2021). Curriculum emphases, mathematics and teaching practices: Swedish upper-secondary physics teachers' views. *International Journal of Science and Mathematics Education*, *19*(3), 499–515. https://doi.org/10.1007/s10763-020-10078-6

Julianto, V., Sumintono, Bambang Almakhi, N. P. Z., Avetazain, H., Wilhelmina, T. M., & Wati, D. A. (2025). Academic motivation scale's psychometric attribute: Analysis using Rasch measurement model. *Current Psychology*, *44*, 114–124. https://doi.org/10.1007/s12144-024-07142-7

Kaniawati, I., Fratiwi, N. J., Danawan, A., Suyana, I., Samsudin, A., & Suhendi, E. (2019). Analyzing students' misconceptions about Newton's laws through Four-Tier Newtonian Test (FTNT). *Journal of Turkish Science Education*, *16*(1), 110–122. https://doi.org/10.12973/tused.10269a

Kiray, S. A., & Simsek, S. (2021). Determination and evaluation of the science teacher candidates' misconceptions about density by using four-tier diagnostic test. *International Journal of Science and Mathematics Education*, *19*(5), 935–955. https://doi.org/10.1007/s10763-020-10087-5

Kurnianto, A., Sunjaya, D. K., Rinawan, F. R., & Hilmanto, D. (2023). Validation and reliability of lifestyle instruments for Indonesian adolescents with hypertension family history: A Rasch model. *Open Access Macedonian Journal of Medical Sciences*, *11*(B), 109–114. https://doi.org/10.3889/oamjms.2023.11127

Kurniawan, F., Samsudin, A., Chandra, D. T., Sriwati, E., Zahran, M., Gani, A. W., Ramadhan, B. P., Aminudin, A. H., & Ramadani, F. (2023). Assessing conceptual understanding of high school students on transverse and stationary waves through Rasch analysis in Malang. *In Journal of Physics: Conference Series*, *2596*(1), 012060. https://doi.org/10.1088/1742-6596/2596/1/012060

Osman, E., BouJaoude, S., & Hamdan, H. (2017). An investigation of Lebanese G7-12 students' misconceptions and fifficulties in genetics and their genetics literacy. *International Journal of Science and Mathematics Education*, *15*(7), 1257–1280. https://doi.org/10.1007/s10763-016-9743-9

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and Quality of Life Outcomes*, *1*, 1–13. https://doi.org/10.1186/1477-7525-1-27

Salimpour, S., Tytler, R., Doig, B., Fitzgerald, M. T., & Eriksson, U. (2022). Conceptualising the Cosmos: Development and validation of the Cosmology concept inventory for high school. *International Journal of Science and Mathematics Education*, *21*, 251–275. https://doi.org/10.1007/s10763-022-10252-y

Samsudin, A., Azizah, N., Fratiwi, N. J., Suhandi, A., Irwandani, I., Nurtanto, M., Yusup, M., Supriyatman, S., Masrifah, M., Aminudin, A. H., & Costu, B. (2024). Development of DIGaKiT: Identifying students' alternative conceptions by Rasch analysis model. *Journal of Education and Learning (EduLearn)*, *18*(1), 128–139. https://doi.org/10.11591/edulearn.v18i1.20970

Samsudin, A., Cahyani, P. B., Purwanto, Rusdiana, D., Efendi, R., Aminudin, A. H., & Coştu, B. (2021). Development of a multitier open-ended work and energy instrument (MOWEI) using Rasch analysis to identify students' misconceptions. *Cypriot Journal of Educational Sciences*, *16*(1), 16–31. https://doi.org/10.18844/cjes.v16i1.5504

Samsudin, A., Rusdiana, D., Efendi, R., Fratiwi, N. J., Aminudin, A. H., & Adimayuda, R. (2021). Development of Predict-Observe-Explain (POE) strategy assisted by Rebuttal texts on Newton's law material with Rasch analysis. *Tadris: Jurnal Keguruan Dan Ilmu Tarbiyah*, *6*(1), 103–115. https://doi.org/10.24042/tadris.v6i1.7641

Samsudin, A., Suhandi, A., Linuwih, S., Kaniawati, I., Fauzi, D. M., Aminudin, A. H., Fratiwi, N. J., & Adimayuda, R. (2023). Preliminary development of Simulation on Refutational-Text (SoRT) to change conception of physics: Rasch and Nvivo analysis. *AIP Conference Proceedings*, 1–7. https://doi.org/10.1063/5.0125826

Samsudin, A., Wulandari, N., Suhandi, A., Yusup, M., Supriyatman, S., Aminudin, A. H., Sumarni, W., & Kapıcı, H. Ö. (2025). Augmented Reality-based Rebuttal Texts (ARaRaT) on momentum-impulse: Rasch analysis on students' conceptual change. *Qubahan Academic Journal*, *5*(1), 368–387. https://doi.org/10.48161/qaj.v5n1a1163

Samsudin, A., Zulfikar, A., Saepuzaman, D., Suhandi, A., Aminudin, A. H., Supriyadi, S., & Coştu, B. (2024). Correcting grade 11 students' misconceptions of the concept of force through the conceptual change model (CCM) with PDEODE*E tasks. *Journal of Turkish Science Education*, *21*(2), 212–231. https://doi.org/10.36681/tused.2024.012

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, *8*, 1–11. https://doi.org/10.1186/1471-2288-8-33

Wati, M., Mahtari, S., Hartini, S., & Amalia, H. (2019). A Rasch model analysis on junior high school students' scientific reasoning ability. *International Journal of Interactive Mobile Technologies*, *13*(7), 141–149. https://doi.org/10.3991/ijim.v13i07.10760

Wei, S., Liu, X., & Jia, Y. (2014). Using Rasch measurement to validate the instrument of Students' Understanding of Models in Science (SUMS). *International Journal of Science and Mathematics Education*, *12*(5), 1067–1082. https://doi.org/10.1007/s10763-013-9459-z

Yang, Y., He, P., & Liu, X. (2018). Validation of an Instrument for Measuring Students' Understanding of interdisciplinary science in grades 4–8 over Multiple Semesters: ARasch measurement study. *International Journal of Science and Mathematics Education*, *16*(4), 639–654. https://doi.org/10.1007/s10763-017-9805-7

Yen, C.-F., Yao, T.-W., & Chiu, Y.-C. (2004). Alternative conceptions in animal classification focusing on Amphibians and Reptiles: A cross-age study. *International Journal of Science and Mathematics Education*, *2*(2), 159–174. https://doi.org/10.1007/s10763-004-1951-z

Zulfa, I., Kusairi, S., Latifah, E., & Jauhariyah, M. N. R. (2019). Analysis of student's conceptual understanding on the work and energy of online hybrid learning. *Journal of Physics: Conference Series*, *1171*(1). https://doi.org/10.1088/1742-6596/1171/1/012045