Infinity Journal of Mathematics Education
Volume 15, No. 3, 2026
p-ISSN 2089-6867
e–ISSN 2460-9285
https://doi.org/10.22460/infinity.v15i3.p921-940

Measuring teachers' beliefs about mathematics teaching: A Rasch analysis of a deep learning approach

Muhtarom[1*], Ahmad Wachidul Kohar[2], Sugiyanti[1], Irkham Ulil Albab[1], Sri Rejeki[3]
[1]Department of Mathematics Education, Universitas PGRI Semarang, Central Java, Indonesia
[2]Department of Mathematics Education, Universitas Negeri Surabaya, East Java, Indonesia
[3]Department of Mathematics, Universiti Pendidikan Sultan Idris, Perak, Malaysia
[*]Correspondence: muhtarom@upgris.ac.id

Received: Apr 29, 2026 | **Revised**: Jul 15, 2026 | **Accepted**: Aug 3, 2026 | **Published Online**: Aug 8, 2026

Abstract

Teachers' beliefs significantly influence the efficacy of mathematics education. The absence of an instrument to assess beliefs about deep learning poses a challenge for evaluating and enhancing mathematics teachers' competence in adopting deep learning. This research aims to develop and test the validity and reliability of an instrument to measure mathematics teachers' beliefs about implementing deep learning. This research involved 108 mathematics teachers from five provinces in Indonesia. The research data were collected using a questionnaire with a three-dimensional scale: mindful learning, meaningful learning, and joyful learning. The collected research data were then analyzed using the Rasch model, assisted by Winstep software version 3.73. The results of the first calibration indicated that the instrument's quality was very good, and the respondents' answers were quite consistent. The analysis was continued by examining each person and item more closely, resulting in the identification of three items as misfits. The second calibration was carried out by eliminating the misfit statement items. In the second calibration, a valid, reliable, and unidimensional instrument was obtained, which did not exhibit bias towards any of the research attributes. The instrument demonstrated very good reliability, and respondents' answers were quite consistent. After repeated calibration, a deep learning belief measurement instrument was obtained, consisting of 17 high-quality items with adequate psychometric properties.

Keywords:
Beliefs, Deep learning, Joyful learning, Meaningful learning, Mindful learning

How to Cite:
Muhtarom, M., Kohar, A. W., Sugiyanti, S., Albab, I. U., & Rejeki, S. (2026). Measuring teachers' beliefs about mathematics teaching: A Rasch analysis of a deep learning approach. *Infinity Journal*, *15*(3), 921-940. https://doi.org/10.22460/infinity.v15i3.p921-940

1. INTRODUCTION

Educational transformation requires a learning approach that integrates technology, promotes flexible learning, and develops 21st-century skills. In this context, deep learning becomes increasingly relevant because it not only focuses on mastering the material but also encourages students to think critically, solve problems, collaborate, be creative, and apply knowledge in various real-life contexts (Feriyanto & Anjariyah, 2024; Hidayat et al., 2025; Meika et al., 2025). Deep learning is an approach that emphasizes the development of meaningful conceptual understanding, critical thinking skills, reflection, and the transfer of information into actual-life scenarios (Darling-Hammond et al., 2019; Perrotta & Selwyn, 2020; Sinha & Ayush, 2018). The implementation of deep learning cannot be separated from teachers' readiness and role as the main facilitators. One of the key factors influencing the success of this implementation is the teacher's belief in the value and effectiveness of the pedagogical approach used (DeJaeghere et al., 2024; Rezaei, 2011). Belief reflects the teacher's perception of their role in creating a meaningful and enjoyable educational setting, and their ability to facilitate the student's overall learning process (Buehl & Beck, 2014; Siswono, Kohar, & Hartono, 2019). The way teachers teach mathematics is not only influenced by their beliefs, but also shapes students' ways of thinking and beliefs about mathematics learning itself (Muhtarom et al., 2020; Siswono, Kohar, Hartono, et al., 2019).

While various dimensions of deep learning have been proposed in previous literature, three key dimensions have been identified through their library study: meaningful learning, mindful learning, and joyful learning (Feriyanto & Anjariyah, 2024; Nafi'ah & Faruq, 2025). Mindful learning encourages students to be fully present in the learning process, conscious of their thoughts and feelings, and to strengthen attention and reflection (Kovač et al., 2023; Roeser et al., 2012). The mindful learning process is shown by the process of students being fully involved, being able to understand learning objectives, being intrinsically motivated to learn, having the ability to consider their educational experiences, and developing learning strategies to achieve goals (Gördesli et al., 2019; Leppma & Darrah, 2024). Teachers who implement mindful learning have positive feelings, are comfortable, and have an optimistic attitude toward achieving learning goals (Ojell et al., 2023; Roeser et al., 2012; Siswanto & Afandi, 2024). To see mindful learning from teachers, it can be identified from the dimensions of interpersonal mindfulness and intrapersonal mindfulness (Frank et al., 2016; Khairani et al., 2021). Intrapersonal mindfulness is the teacher's ability to comprehend themselves, and interpersonal mindfulness is the teacher's ability to understand, recognise, and react to students' feelings, needs, and behavior in the context of social relationships (Khairani et al., 2021).

Meaningful learning emphasizes the relationship between teaching materials and students' real-life experiences (Feriyanto & Anjariyah, 2024; Koh, 2013; Lee, 2018). Meaningful learning emphasizes understanding how the information learned is interrelated (Anggoro et al., 2024; George & Sanders, 2017; Ghazali et al., 2020; Hsbollah & Hassan, 2022). The learning process is not only limited to mastering the content of mathematical materials, but is more oriented towards the ability to link new concepts with the knowledge that students already possess and apply them in various real-world contexts. Meaningful learning contrasts with rote learning, which involves memorizing information repeatedly (Raihan & Han, 2012). Meaningful learning involves understanding how learning complements each other, whereas rote learning involves memorizing replicated knowledge. To measure meaningful learning from teachers, it can be identified from the dimensions of active learning, cooperative learning, constructive learning, authentic learning, and intentional learning (Ghazali et al., 2020). In the learning process, cooperative learning places a strong emphasis on teacher-student interaction (George & Sanders, 2017). Active learning occurs when students participate in group discussions, demonstrations, presentations, and instructor-led activities, in addition to listening to teachers (Feriyanto & Anjariyah, 2024). Authentic learning connects learning activities that involve students in real-world experiences, recognizing problems, and solving mathematical problems. Constructive learning emphasizes students developing mathematical understanding through the integration of new knowledge and prior knowledge (Hsbollah & Hassan, 2022). To develop mathematical comprehension, students must synthesize new ideas, relate them to existing knowledge, and provide clear explanations. Students who actively manage their own learning and establish learning objectives are said to be engaging in intentional learning. Students plan their daily learning tasks and find ways to solve the problems they encounter (George & Sanders, 2017; Ghazali et al., 2020).

Joyful learning increases students' intrinsic motivation, creativity, and psychological well-being by creating a fun learning environment (Anggoro et al., 2024; Ardiyanti et al., 2021; Feriyanto & Anjariyah, 2024; Mudrikah et al., 2025). Enthusiastic learning fosters emotional connections in pupils, which facilitates comprehension, memory, and application of information. Students will be intrinsically motivated to be more intelligent, creative, and engaged in class when they find the learning process enjoyable (Mooduto & Musa, 2021; Sulistyowati et al., 2024; Ulfa & Hidayat, 2024). Joyful learning creates a memorable learning experience (Mooduto & Musa, 2021). Students who feel comfortable and have their needs met, such as a sense of security, belonging, affection, appreciation, and self-actualization, will make learning more enjoyable.

These three dimensions are important in deep learning practices. Teachers' classroom practices are influenced by their beliefs (Muhtarom et al., 2020; Siswono, Kohar, & Hartono, 2019). If teachers believe that traditional methods are effective for teaching mathematics, they are more likely to use them. This demonstrates the importance of identifying a teacher's beliefs early on. However, there is currently no instrument specifically available to measure teachers' beliefs in implementing integrated deep learning. Existing instruments generally only measure teacher beliefs in learning in general (Muhtarom, 2024; Purnomo, 2017). The absence of an instrument to measure the dimensions of deep learning beliefs affects the difficulty of evaluating and improving teacher readiness in integrating deep learning. Based on this background, the study’s objective is to develop and test the validity and reliability of an instrument for measuring mathematics teacher beliefs in implementing deep learning. The formulation of the research problem is: How valid and reliable is the instrument for measuring mathematics teacher beliefs in implementing deep learning?

2. METHOD

This type of research is development research (Nieveen & Folmer, 2013). This research aims to develop an instrument to measure teacher beliefs in implementing deep learning. Development research focuses on the creation and validation of instruments that can be used to measure a variable, in this case, teacher beliefs in deep mathematics learning. The instrument developed in this research is expected to provide accurate information about teachers' implementation of the deep learning approach.

2.1. Respondents

The respondents in this research were mathematics teachers who voluntarily participated by completing a research questionnaire. Their participation was based on their informed consent, and they filled out the questionnaire based on their circumstances and experiences. This research involved 108 mathematics teachers from five provinces in Indonesia. A total of 9 teacher respondents came from East Java Province, 76 teacher respondents came from Central Java Province, 11 teacher respondents came from West Java Province, 8 teacher respondents came from Aceh Province, and 4 teacher respondents came from South Sulawesi Province. A total of 32 mathematics teacher respondents were male, and the rest were female. Table 1 shows the characteristics of the research participants. This sample size is adequate for analysis using Rasch modeling (O'Neill et al., 2020).

Table 1. Characteristics of research respondents

Attribute	Category	Code	Person (n)	(%)
Province	East Java	1	9	8.33
	Central Java	2	76	70.37
	West Java	3	11	10.19
	Aceh	4	8	7.41
	South Sulawesi	5	4	3.70
Gender	Male	L	32	29.63
	Female	P	76	70.37
Age	21 – 30 years	A	58	53.70
	31 – 40 years	B	20	18.52
	41 – 50 years	C	22	20.37
	51 – 60 years	D	4	7.41
School Level	Junior High School	G	66	61.11
	Senior High School	A	42	38.89
Number of students in the school	Small school (up to 300 students)	K	38	35.19
	Medium school (between 300 and 1000 students)	S	51	47.22
	Large school (more than 1000 students)	B	19	17.59
School Origin Area	Rural	D	30	27.78
	Small towns and semi-densely populated areas	K	56	51.85
	Urban	B	22	20.37

Participants in this research were coded according to Table 1. The first three digits represent the respondent's serial number, followed by the 4th to 9th characters, which are the attributes of the research respondents, including province, gender, age, school level, number of students, and area of origin of the school. For example, a person with code 0392PAGSB can be explained that the respondent's serial number is 039, comes from Central Java province, is female, aged 21-30 years, teaches in junior high school, the number of students in the school is between 300 and 1000 students, and school in urban area. Respondent attributes are used in differential item function analysis (Julianto et al., 2025).

2.2. Data Collection

Research data on teachers' beliefs about implementing deep learning were collected using a questionnaire with a scale comprising three main dimensions: meaningful learning, mindful learning, and joyful learning. Table 2 shows that each dimension is described in several sub-dimensions, which are then developed into statement items. The research respondents answered the statements to determine teachers' beliefs about implementing deep learning. The questionnaire consisted of 10 favourable statements and 10 unfavourable statements. Each item uses a 4-point Likert scale. Respondents are required to read each statement item more attentively and not answer haphazardly when unfavorable items are present (Vigil-Colet et al., 2020; Weijters et al., 2013).

Table 2. Dimension deep learning

Dimension	Sub Dimension	Description	Favourable	Unfavourable
Mindful learning	Intrapersonal mindfulness	Ability to understand oneself	1	6
	Interpersonal mindfulness	The ability to understand and respond to the feelings, needs, and behaviours of others in the context of social relationships	7	2
Meaningful learning	Cooperative learning	Emphasises interaction between students and teachers in the learning process	3	8
	Active learning	Involves students in learning activities	9	4
	Authentic learning	Involves students in real-world experiences, recognising and solving problems from experience	10	5
	Constructive learning	Students develop understanding through the integration of new knowledge and prior knowledge	16	11
	Intentional learning	Students' ability to set their own learning goals, regulate learning, identify gaps in understanding, and address any lack of understanding found	12	17
Joyful learning	Learning environment	Creating a learning environment so that students feel comfortable and want to be involved in the learning process	18	13
	Increasing motivation	The learning process is directed to increase	14	19

Dimension	Sub Dimension	Description	Favourable	Unfavourable
		curiosity, enthusiasm, creativity, and active involvement		
	Student rewards	Giving rewards in the form of verbal praise, badges, points, or other appropriate forms	20	15

2.3. Data Analysis

To obtain valid and reliable instruments, testing was carried out using Rasch modeling (Andrich & Marais, 2019; Boone et al., 2014; Chan et al., 2021; Medvedev & Krägeloh, 2022; Sumintono, 2018). Rasch model analysis was chosen because it is able to estimate response probabilities based on the interaction between respondents and item characteristics, thus producing more objective and accurate measurements. Winsteps software version 3.73 was then used to evaluate the gathered research data. The following are the measurements carried out in the Rasch model analysis, namely: (1) item fit model, (2) person fit model, (3) rating scale analysis, (4) dimensionality, (5) Differential Item Functioning (DIF), and (6) Cronbach's Alpha. Item fit order is used to check item misfits, and person fit order is used to check person misfits (Andrich & Marais, 2019; Medvedev & Krägeloh, 2022). Rating scale analysis is used to determine whether each rating option is confusing for respondents (Akhtar & Sumintono, 2023; Ling Lee et al., 2021; Setiawan et al., 2018; Wei Chan et al., 2015). Dimensionality testing is used to evaluate whether the instrument developed by the researcher accurately measures what it is intended to measure (Adi et al., 2022; Muhtarom, 2024; Setiawan et al., 2018). Differential Item Functioning (DIF) testing is conducted to ensure that the instrument developed is unbiased (Medvedev & Krägeloh, 2022). Meanwhile, reliability analysis is conducted by considering the Cronbach's alpha value to assess the overall reliability of the interaction between statement items and respondents (Adi et al., 2022; Othman et al., 2012). All criteria in the data analysis of this research are presented in Table 3.

3. RESULTS AND DISCUSSION

3.1. Results

Data calibration was performed using Rasch modelling with Winstep 3.73 software to eliminate misfit items and persons. The findings from the initial calibration indicated that the instrument's quality was quite good and that the respondents' responses were highly stable. The analysis was continued by examining each person and item more closely, resulting in the identification of three items as misfit due to the Differential Item Functioning (DIF) factor of respondent age: item 9, DIF by school level: item 17, and DIF by school area of origin: item 5. This bias indicates that these items are interpreted differently across groups. Items affected by DIF should be removed because they may introduce bias against certain respondent groups, thereby no longer fairly representing the construct being measured. For example, item 5 is understood differently due to the uneven distribution of learning environments between urban and rural schools. The second calibration was conducted by excluding the misfit statement items. The second calibration obtained very good instrument quality, and the respondents' answers were quite consistent. The analysis was further refined by examining each person and item in greater detail. The results showed that all persons and items were a good fit.

3.1.1. Model Fit: Item Summary

This analysis aims to evaluate the instrument items' fit to the Rasch model using fit statistics. A total of 17 items were deemed fit out of 20 items developed by the researcher (see Appendix). Figure 1 shows that the Outfit MNSQ and Infit MNSQ values are 0.99 and 0.98, respectively. This value is very good because it is close to the ideal value of 1.00. For Outfit ZSTD and Infit ZSTD, respectively, are -0.00 and -0.10. Due to its proximity to the ideal value of 0.00, this value is very good. The calibration results also indicate that the items in the scale have very good measurement accuracy, as indicated by an SEM of 0.42 (less than 0.5) (Setiawan et al., 2018; Wei Chan et al., 2015).

```
     SUMMARY OF 108 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      50.1      17.0       1.16     .42       1.03     .0    .99    -.1 |
| S.D.       4.2        .0        .84     .05        .44    1.2    .36    1.1 |
| MAX.      66.0      17.0       5.66     .82       3.16    3.5   1.97    2.7 |
| MIN.      42.0      17.0       -.16     .39        .36   -2.6    .38   -2.3 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .47 TRUE SD     .70  SEPARATION  1.49  Person RELIABILITY  .69 |
|MODEL RMSE    .42 TRUE SD     .72  SEPARATION  1.72  Person RELIABILITY  .75 |
| S.E. OF Person MEAN = .08                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = .97
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .63

     SUMMARY OF 17 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                         MODEL         INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN     318.1     108.0        .00     .17        .98    -.1    .99     .0 |
| S.D.      61.7        .0       1.66     .02        .13    1.2    .15    1.1 |
| MAX.     386.0     108.0       2.86     .19       1.27    2.2   1.29    2.3 |
| MIN.     197.0     108.0      -1.98     .14        .73   -2.9    .71   -2.0 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .17 TRUE SD    1.65  SEPARATION  9.62  Item   RELIABILITY  .99 |
|MODEL RMSE    .17 TRUE SD    1.65  SEPARATION  9.76  Item   RELIABILITY  .99 |
| S.E. OF Item MEAN = .42                                                     |
-------------------------------------------------------------------------------
UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -1.00
1836 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 3088.65 with 1710 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): .5922
```

Figure 1. Item summary

The item reliability value of the teacher belief measurement instrument in the implementation of deep learning is 0.99. This shows that the instrument developed is reliable, with a perfect category (more than 0.94). This value indicates that the sample is sufficient to confirm the level of difficulty in measuring beliefs (Adi et al., 2022; Muhtarom, 2024). This is supported by the grouping of items calculated from the separation value. The item separation value is 9.62; then $H = ((4 \times 9.62) + 1) / 3 = 13.16$, which means there are 13 item groups. To check for fit and misfit items, the Outfit MNSQ value is used. The Outfit MNSQ value accepted is $0.5 < MNSQ < 1.5$. Another approach is to examine the Pt Mean Corr value. The criteria used are Pt Measure Corr > 0.00 (Akhtar & Sumintono, 2023; Julianto et al., 2025). As shown in Figure 2, all items fall within the required range of values. This means that all items of the instrument measuring teachers' beliefs in implementing deep learning are declared fit.

Person: REAL SEP.: 1.49 REL.: .69 ... Item: REAL SEP.: 9.62 REL.: .99

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
7	249	108	1.87	.14	1.27	2.2	1.29	2.3	A .48	.44	41.7	49.1	Item8
5	260	108	1.66	.14	1.17	1.4	1.22	1.7	B .50	.43	50.9	50.5	Item6
2	260	108	1.66	.14	1.12	1.0	1.18	1.4	C .47	.43	39.8	50.5	Item2
13	197	108	2.86	.14	1.01	.2	1.08	.7	D .23	.47	52.8	47.7	Item15
11	271	108	1.43	.14	1.03	.3	1.06	.5	E .39	.42	54.6	53.3	Item13
1	386	108	-1.98	.19	1.05	.5	1.00	.1	F .21	.28	57.4	62.1	Item1
9	242	108	2.01	.14	1.02	.3	1.03	.3	G .38	.44	41.7	47.9	Item11
8	358	108	-1.00	.18	.99	.0	.98	-.1	H .26	.33	62.0	64.0	Item10
4	340	108	-.41	.18	.99	.0	.94	-.3	I .53	.35	73.1	68.5	Item4
16	257	108	1.72	.14	.94	-.5	.98	-.1	h .35	.43	52.8	50.5	Item19
14	385	108	-1.94	.19	.98	-.1	.96	-.2	g .33	.28	65.7	62.0	Item16
10	386	108	-1.98	.19	.95	-.4	.92	-.4	f .31	.28	66.7	62.1	Item12
6	372	108	-1.47	.19	.92	-.7	.92	-.5	e .27	.31	66.7	61.0	Item7
3	369	108	-1.37	.18	.92	-.7	.89	-.7	d .39	.31	68.5	61.5	Item3
12	344	108	-.54	.18	.87	-.9	.86	-.9	c .29	.35	70.4	67.8	Item14
15	351	108	-.77	.18	.73	-2.1	.72	-1.9	b .39	.34	74.1	65.9	Item18
17	380	108	-1.76	.19	.73	-2.9	.71	-2.0	a .47	.29	75.0	61.0	Item20
MEAN	318.1	108.0	.00	.17	.98	-.1	.99	.0			59.6	58.0	
S.D.	61.7	.0	1.66	.02	.13	1.2	.15	1.1			11.4	7.1	

Figure 2. Item misfit order

3.1.2. Model Fit: Person Summary

This analysis aims to evaluate the suitability of each respondent's response pattern to the Rasch model, so that consistent responses and responses that deviate from the measurement pattern are obtained. As shown in Figure 1, the Outfit MNSQ and Infit MNSQ values are 0.99 and 1.03, respectively. This value is very good because it is close to the ideal value of 1.00. For Outfit ZSTD and Infit ZSTD, respectively, are -0.10 and 0.00. This value is very good because it is close to the ideal value of 0.00. The calibration results also show that the teachers who are respondents in the study have perfect measurement accuracy, as indicated by the SEM value of 0.08 (less than 0.25) (Setiawan et al., 2018; Wei Chan et al., 2015).

Table 3. Summary of Rasch model analysis results

Analysis	Parameter	Criteria	Value	Category
Model Fit: Item Summary	Measure item	Poor: SEM > 2 Fair: SEM 1-2 Good: SEM 1-0.5 Very good: SEM 0.5-0.25 Perfect: SEM < 0.25	M: 0.00 SD: 1.66 SEM: 0.42	Very Good
	Item measurement reliability	Poor < 0.67 Sufficient 0.67-0.80 Good 0.81-0.90 Very good 0.91-0.94 Perfect > 0.94	0.99	Perfect
	Strata item separation (H) =	Fair 2-3 Good 3-4	13.16	Perfect

Analysis	Parameter	Criteria	Value	Category
	[(4×item separation index) + 1]/3	Very good 4-5 Perfect >5		
	Outfit MNSQ item	0.50-1.50	0.5 – 1.50	Suitable
	Point measurement correlation	≥ 0.00	0.21 – 0.51	Suitable
Model Fit: Person Summary	Measure person	Poor SEM>2 Fair SEM 1-2 Good SEM 1-0.5 Very good SEM 0.5-0.25 Perfect SEM <0.25)	M: 1.16 SD: 0.84 SEM: 0.08	Perfect
	Person measurement reliability	Poor <0.67 Sufficient 0.67-0.80 Good 0.81-0.90 Very good 0.91-0.94 Perfect >0.94	0.69	Fair
	Strata item separation (H) = [(4×item separation index) + 1]/3	Fair 2-3 Good 3-4 Very good 4-5 Perfect >5	2.32	Fair
	Outfit MNSQ item	0.00-2.00	0.38 – 1.97	Suitable
	Point measurement correlation	≥ 0.00	0.31 – 0.94	Suitable
Rating Scale Analysis	Responses per category	≥10	1: 133 2: 334 3: 870 4: 449	Suitable
	Measurement means	Increases monotonically between rating scales	1: -1.14 2: -0.78 3: 1.28 4: 2.85	Suitable
	Outfit MNSQ	0.00-2.00 logits	1: 1.26 2: 0.79 3: 0.92 4: 1.01	Suitable
	Andrich threshold	Increases monotonically between rating scales	1: NONE 2: -1.94 3: -0.73 4: 2.68	Suitable
	Adjusted threshold distance	1.4 - 5.0	Category 1-2 is -3.22 to -1.36 = 1.86;	Suitable

Analysis	Parameter	Criteria	Value	Category
			Category 2-3 is -1.36 to 1.02 = 2.38; Category 3-4 is 1.02 to 3.80 = 2.78	
Dimen-sionality	Raw variance in data explained by measurements	Poor <20% Fair 20%-40% Good 40%-60% Very good >60%	52.6%	Good
	Unexplained variance in the 1st-5th contrast of residual PCA	Poor >15.0% Fair 10.0%-15.0% Good 5.0%-10.0% Very good 3.0%-5.0 Excellent <3.0%)	1st contrast = 10.5% 2nd contrast = 5.6% 3rd contrast = 3.8% 4th contrast = 3.6% 5th contrast = 3.0%	Fair Good Very Good Very Good Very Good
DIF	Items with significant DIF	p <0.05	There are no misfit items	Suitable
Reliability	Cronbach's Alpha	Very poor <0.50 Poor, 0.50-0.60 Fair: 0.60-0.70 Good 0.70-0.80 Excellent > 0.80	0.63	Fair

The person reliability value in measuring teacher beliefs in implementing deep learning is 0.69, with a fairly reliable category (Chan et al., 2021; Julianto et al., 2025). The separation value is used to distinguish between different levels of a person's abilities. The item separation value is 1.49, and H = 2.32, which indicates that there are three categories of respondents' beliefs regarding the implementation of deep learning. To check fit and misfit items, the Outfit MNSQ value is used. The accepted Outfit MNSQ range is 0.00 < MNSQ < 2.00. Another approach is to examine the Pt Mean Corr value. The criteria used are Pt Measure Corr> 0.00 (Akhtar & Sumintono, 2023; Julianto et al., 2025). Table 3 shows that the Outfit MNSQ and Pt Mean Corr values for all persons are within the required range. This shows that the respondents in this research were declared fit because the answer patterns given by the respondents were in accordance with the expectations of the Rasch model, so that the resulting responses were considered consistent and reflected the characteristics measured by the instrument.

3.1.3. Rating Scale Analysis

The rating scale analysis presented in Table 3 indicates that each rating option does not confuse respondents. The average measurement results for all categories increase in a monotonic manner. The average measurement starting from category 1 is -1.14, category 2 is

-0.78, category 3 is 1.28, and category 4 is 2.85. The Andrich Threshold calibration is another recommended metric that determines whether the polytomy being utilized is adequate (Medvedev & Krägeloh, 2022). Figure 3 clearly shows that the Andrich Threshold value moves from category 1, with a value of NONE, to category 2, with a value of -1.94, to category 3, with a value of -0.73, and finally to category 4, with a value of 2.68. This demonstrates that responders may choose from the provided answer possibilities. Furthermore, it is also known how far apart each rating is. With a difference of 1.86, categories 1-2 fall between -3.22 and -1.36; categories 2-3 begin at -1.36 and 1.02, with a distance of 2.38; while categories 3-4 begin at 1.02 and 3.80, with a distance of 2.78. It can be seen that the adjusted threshold distance meets the criteria in the interval 1.4-5.0 (Ling Lee et al., 2021; Wei Chan et al., 2015).

```
SUMMARY OF CATEGORY STRUCTURE.  Model="R"
-------------------------------------------------------------------
|CATEGORY   OBSERVED|OBSVD SAMPLE|INFIT OUTFIT|| ANDRICH |CATEGORY|
|LABEL SCORE COUNT %|AVRGE EXPECT|  MNSQ  MNSQ||THRESHOLD| MEASURE|
|-------------------+------------+------------++---------+--------|
|  1   1      133   7| -1.14 -1.35|  1.19  1.26||  NONE   |( -3.22)| 1
|  2   2      334  18|  -.78  -.61|   .83   .79||  -1.94  |  -1.36 | 2
|  3   3      870  47|  1.28  1.25|   .96   .92||   -.73  |   1.02 | 3
|  4   4      499  27|  2.85  2.85|  1.01  1.01||   2.68  |(  3.80)| 4
-------------------------------------------------------------------
```

Figure 3. Rating scale analysis

3.1.4. Dimensionality

Dimensionality analysis is carried out to ensure that the instrument measures a single construct, thereby accounting for its validity. Dimensionality analysis reveals that the overall variation in observations accounts for 52.6%, with a notable proportion falling into the good category. Figure 4 shows that the unexplained variance for the first contrast is 10.5%, the second is 5.6%, the third is 3.8%, the fourth is 3.6%, and the fifth is 3.0%. This shows that this instrument can effectively measure mathematics teachers' beliefs in deep learning (see Table 3). As expected, all items measure one dimension, so the unidimensionality requirement has been met (Akhtar & Sumintono, 2023; Andrich & Marais, 2019; Chan et al., 2021; Ling Lee et al., 2021).

```
     Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                            -- Empirical --    Modeled
Total raw variance in observations     =    35.9 100.0%         100.0%
  Raw variance explained by measures   =    18.9  52.6%          52.9%
    Raw variance explained by persons  =     2.6   7.2%           7.3%
    Raw Variance explained by items    =    16.3  45.4%          45.7%
  Raw unexplained variance (total)     =    17.0  47.4% 100.0%   47.1%
    Unexplned variance in 1st contrast =     3.8  10.5%  22.2%
    Unexplned variance in 2nd contrast =     2.0   5.6%  11.9%
    Unexplned variance in 3rd contrast =     1.4   3.8%   8.0%
    Unexplned variance in 4th contrast =     1.3   3.6%   7.5%
    Unexplned variance in 5th contrast =     1.1   3.0%   6.3%
```

Figure 4. Output dimensionality

3.1.5. Differential Item Functioning (DIF)

This analysis was conducted to ensure that there were no instrument items that were biased toward certain respondent groups, so that each item measured the construct fairly and equally. DIF investigates whether instrument items function differently for different groups of respondents (Sumintono, 2018). In this research, several attributes of the research respondents are examined, including province, gender, age, school level, number of students, and the school's area of origin, to investigate the DIF of the instrument items. An item is said to have significant bias if its p-value is less than 0.05. In the first calibration stage, three instrument items were identified as exhibiting DIF (item bias): DIF by respondent age (item 9), DIF by school level (item 17), and DIF by school area of origin (item 5). During the second calibration, it was determined that none of the instrument items exhibited bias against any research attribute (see Table 3). For example, Figure 5 presents the DIF analysis results by respondents' area demographics, divided into three groups: rural areas, small towns, and large cities.

DIF class specification is: DIF=$S9W1

Person CLASSES	SUMMARY DIF CHI-SQUARE	D.F.	PROB.	BETWEEN-CLASS MEAN-SQUARE	t=ZSTD	Item Number	Name
3	2.5866	2	.2709	.3177	-.6197	1	Item1
3	.2604	2	.8793	.0315	-1.7196	2	Item2
3	3.0280	2	.2169	.5855	-.1570	3	Item3
3	4.5414	2	.1013	.5889	-.1521	4	Item4
3	1.9369	2	.3760	.2376	-.8085	5	Item6
3	1.6754	2	.4292	.2034	-.9023	6	Item7
3	.4107	2	.8147	.0668	-1.4496	7	Item8
3	1.6324	2	.4386	.2363	-.8118	8	Item10
3	.9095	2	.6325	.1248	-1.1674	9	Item11
3	.0174	2	.9922	.0027	-2.2473	10	Item12
3	.7029	2	.7024	.0957	-1.2944	11	Item13
3	.5272	2	.7680	.0990	-1.2789	12	Item14
3	.4988	2	.7791	.0597	-1.4944	13	Item15
3	1.1152	2	.5698	.2112	-.8802	14	Item16
3	2.5620	2	.2743	.3225	-.6095	15	Item18
3	.6000	2	.7401	.0709	-1.4248	16	Item19
3	2.8804	2	.2337	.4524	-.3636	17	Item20

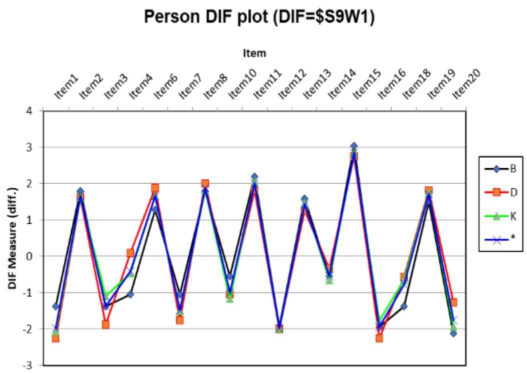

Figure 5. DIF based on respondent demographics

3.1.6. Cronbach's Alpha Reliability

Cronbach's Alpha value indicates the level of internal consistency between instrument items and respondent responses, thus reflecting the instrument's overall reliability. A Cronbach's Alpha value close to 1 indicates excellent internal measurement consistency. The results of the analysis obtained a Cronbach Alpha value of 0.63, with a fairly good interpretation (Adi et al., 2022; Othman et al., 2012). This shows that the interaction between 108 research respondents and 17 items of the mathematics teacher belief measurement instrument in deep learning is reliable.

3.2. Discussion

The deep learning belief measurement instrument was developed with three main dimensions, namely mindful, meaningful, and joyful learning. The use of Rasch modelling as an instrument analysis method was chosen because of its ability to test construct validity, reliability, and data suitability to the expected theoretical model (Andrich & Marais, 2019; Boone et al., 2014; Ling Lee et al., 2021; Setiawan et al., 2018). The results of the Rasch analysis provide a comprehensive picture of the quality of the items, measurement scales, and internal consistency of the instruments developed. The statistical fit model analysis shows that the majority of items in the instrument have infit and outfit mean square (MNSQ) values within the acceptable range. This indicates that the items are empirically in line with the Rasch model and can measure the construct of teacher beliefs validly (Akhtar & Sumintono, 2023; Chan et al., 2021; Julianto et al., 2025; Ling Lee et al., 2021; Wei Chan et al., 2015). This finding aligns with previous research that developed a Rasch-based teacher belief instrument, emphasizing the importance of item validity in obtaining accurate measurements in an educational context (Muhtarom, 2024).

An item reliability of 0.99 indicates very high reliability, suggesting that the items in this instrument consistently measure the same construct (Chan et al., 2021; Julianto et al., 2025). The item separation value of 9.62 also strengthens the instrument's ability to distinguish item difficulty levels, which in turn increases construct validity. This is important, given that a good instrument not only measures beliefs in general but also captures variations in belief intensity among respondents. Although person reliability is lower than item reliability, the results still fall within the fairly reliable category (Adi et al., 2022; Andrich & Marais, 2019; Julianto et al., 2025; Othman et al., 2012). These results indicate that the instrument can distinguish among three distinct categories of teacher beliefs. Thus, the respondents in this research can be considered representative and appropriate for testing the belief instrument for deep learning.

The developed instrument successfully met the unidimensionality requirements with a raw observation variance of 52.6%, which is categorized as good for social and educational research (Akhtar & Sumintono, 2023; Andrich & Marais, 2019; Chan et al., 2021; Ling Lee et al., 2021). The low first-to-fifth contrasts indicate that the unexplained variance is relatively small, suggesting that this instrument effectively measures a single main construct: teachers' beliefs in implementing deep learning. The instrument has a consistent internal structure and does not include other latent constructs that could interfere with measurement. These results strengthen the evidence that all items in the instrument work harmoniously in measuring the same construct. The analysis results show that the three aspects (mindful learning, meaningful learning, and joyful learning) used in the development of the instrument form a construct that together reflects a single latent construct, namely teachers’ beliefs in the implementation of deep learning. Thus, these three aspects can be viewed as complementary conceptual dimensions for explaining teachers’ beliefs, while, empirically, all items still meet the unidimensionality assumption of the Rasch model. These findings indicate that the developed instrument has good construct validity and is suitable for measuring teachers' beliefs regarding the implementation of deep learning.

The person reliability value and Cronbach's Alpha value are still quite acceptable in the context of initial instrument development and a relatively limited number of items (Adi et al., 2022; Othman et al., 2012). This value indicates adequate internal consistency between the items and the constructs being measured, but also indicates improvements to the deep learning belief measurement instrument. Improvements can be made by adding or modifying item variations to increase reliability, and it is highly recommended to include a wider, and more diversified sample.

The application of Rasch modeling as an analytical technique was essential to the success of this investigation. Rasch modeling serves as a powerful statistical instrument, providing numerous benefits in educational research, especially in the validation of measurement tools. It examines construct validity, confirming that the instrument appropriately represents the theoretical characteristics of deep learning, and also assesses the dependability of the instrument across various sample populations. Furthermore, Rasch modeling facilitates the assessment of data appropriateness, confirming that the data aligns with the anticipated theoretical model, which is essential for generating significant and dependable outcomes (Andrich & Marais, 2019; Boone et al., 2014; Ling Lee et al., 2021). The findings of this research validate that the developed scale is capable of measuring mathematics teachers' beliefs about the implementation of deep learning through three main dimensions. The dimensions of mindful learning reflect teachers' beliefs about encouraging students' cognitive engagement, reflection, and self-regulation across learning stages. The meaningful learning dimension demonstrates the importance of connecting mathematical ideas to real-life contexts, culture, and students' future needs to make learning more relevant. Meanwhile, the joyful learning dimension emphasizes that a positive, safe, and motivating learning environment is crucial for increasing student engagement and learning success. The findings of this study can contribute to the achievement of Sustainable Development Goal (SDG) 4, especially the target of providing inclusive, equitable, quality education and encouraging lifelong learning opportunities. This validated measurement tool can be used to map teachers' beliefs about deep learning as a basis for developing professional competency improvement programs. The measurement results can also be used to identify aspects of beliefs that still need strengthening, leading to more targeted teacher training, particularly to expand student-centered, contextual, reflective, and enjoyable mathematics learning.

A notable strength of the current research is the utilization of Rasch modeling to examine the reliability and validity of the belief assessment instrument. This method not only validates the instrument's theoretical strength but also provides a detailed analysis of the data, facilitating the identification of areas for future improvement. The Rasch model's ability to examine the alignment of individual data points with the underlying theoretical framework makes it an essential instrument for ensuring the accuracy and validity of educational assessment tools. This research uses the Rasch model for instrument analysis, offering a robust framework for the development and validation of educational assessment tools. Subsequent research should continue to investigate the use of this instrument across various educational settings, analyzing the evolution of instructors' views and their influence on the implementation of deep learning to fulfil the overarching objectives of SDG 4.

4. CONCLUSION

The calibrated instrument makes a significant contribution to measuring teachers' beliefs about deep learning. The test results show that this instrument is valid, reliable, unidimensional, and free from bias. These findings strengthen the argument that the development of a Rasch-based teacher belief instrument not only provides a valid and reliable measurement instrument but also yields diagnostic data to identify the level of teachers' beliefs, which can ultimately support the improvement of learning quality in schools. This study significantly advances psychometric tools in instruction, opening opportunities for further research to enhance the reliability and generalizability of results with larger, more diverse samples.

Acknowledgments

The authors are grateful to DPPM Kemendiktisaintek for granting the Fundamental research grants under contract 127/C3/DT.04.00/PL/2025. Additionally, the authors express their gratitude to all teachers and students who were willing to participate as research respondents.

Declarations

Author Contribution	: M: Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, and Writing - review & editing; AWK: Writing - review & editing; S: Methodology, and Writing - original draft; IUA: Project administration, Supervision, and Validation; SR: Writing - review & editing.
Funding Statement	: This research was funded by DPPM Kemendiktisaintek for granting the Fundamental research grants.
Conflict of Interest	: The authors declare no conflict of interest.
Additional Information	: Additional information is available for this paper.

REFERENCES

Adi, N. R. M., Amaruddin, H., Adi, H. M. M., & A'yun, I. L. Q. (2022). Validity and reliability analysis using the rasch model to measure the quality of mathematics test items of vocational high schools. *Journal of Research and Educational Research Evaluation*, *11*(2), 103–113.

Akhtar, H., & Sumintono, B. (2023). Rasch analysis of the international personality item pool big five markers questionnaire: Is longer better? *Primenjena psihologija*, *16*(1), 3–28. https://doi.org/10.19090/pp.v16i1.2401

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: Measuring in the educational, social and health sciences*. Springer Singapore. https://doi.org/10.1007/978-981-13-7496-8

Anggoro, S., Fitriati, A., Thoe, N. K., Talib, C. A., & Mareza, L. (2024). Differentiated instruction based on multiple intelligences as promising joyful and meaningful learning. *International Journal of Evaluation and Research in Education (IJERE)*, *13*(2), 1194–1204. https://doi.org/10.11591/ijere.v13i2.24791

Ardiyanti, S., Qurbaniah, M., & Muldayanti, N. D. (2021). Joyful learning model: Improving higher order thinking skill and students' learning motivation at senior high school. In *Proceedings of KOBI 2nd International Conference on Management of Tropical Biodiversity for Human Welfare: From Ecosystem to Molecular*, (Vol. 1, pp. 33–40).

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer Dordrecht. https://doi.org/10.1007/978-94-007-6857-4

Buehl, M. M., & Beck, J. S. (2014). The relationship between teachers' beliefs and teachers' practices. In M. G. Gill & H. Fives (Eds.), *International handbook of research on teachers' beliefs* (pp. 66–84). Routledge. https://doi.org/10.4324/9780203108437-6

Chan, S.-W., Looi, C.-K., & Sumintono, B. (2021). Assessing computational thinking abilities among Singapore secondary students: A Rasch model measurement analysis. *Journal of Computers in Education*, *8*(2), 213–236. https://doi.org/10.1007/s40692-020-00177-2

Darling-Hammond, L., Oakes, J., Wojcikiewicz, S., Hyler, M. E., Guha, R., Podolsky, A., Kini, T., Cook-Harvey, C., Mercer, C., & Harrell, A. (2019). *Preparing teachers for deeper learning*. Learning Policy Institute.

DeJaeghere, J., Dao, V., & Nguyen, T. (2024). Education systems change: cultural beliefs and practices that support and inhibit deep learning in Vietnam. *Journal of Educational Change*, *25*(3), 555–577. https://doi.org/10.1007/s10833-024-09505-0

Feriyanto, F., & Anjariyah, D. (2024). Deep learning approach through meaningful, mindful, and joyful learning: A library research. *Electronic Journal of Education, Social Economics and Technology*, *5*(2), 208–212. https://doi.org/10.33122/ejeset.v5i2.321

Frank, J. L., Jennings, P. A., & Greenberg, M. T. (2016). Validation of the mindfulness in teaching scale. *Mindfulness*, *7*(1), 155–163. https://doi.org/10.1007/s12671-015-0461-0

George, A., & Sanders, M. (2017). Evaluating the potential of teacher-designed technology-based tasks for meaningful learning: Identifying needs for professional development. *Education and Information Technologies*, *22*(6), 2871–2895. https://doi.org/10.1007/s10639-017-9609-y

Ghazali, N., Mustakim, S. S., & Nordin, M. S. (2020). Development of meaningful learning scale (MeLearn). *Development*, *7*(9), 2107–2117.

Gördesli, M. A., Arslan, R., Çekici, F., Sünbül, Z. A., & Malkoç, A. (2019). The psychometric properties of the mindfulness in teaching scale in a Turkish sample. *Universal Journal of Educational Research*, *7*(2), 381–386. https://doi.org/10.13189/ujer.2019.070210

Hidayat, W., Aripin, U., & Widodo, S. A. (2025). Integration of ethno-modelling and 3N: An innovative digital worksheet framework to enhance students' mathematical critical thinking skills. *Infinity Journal*, *14*(4), 1019–1042. https://doi.org/10.22460/infinity.v14i4.p1019-1042

Hsbollah, H. M., & Hassan, H. (2022). Creating meaningful learning experiences with active, fun, and technology elements in the problem-based learning approach and its implications. *Malaysian Journal of Learning and Instruction*, *19*(1), 147–181. https://doi.org/10.32890/mjli2022.19.1.6

Julianto, V., Sumintono, B., 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*(1), 114–124. https://doi.org/10.1007/s12144-024-07142-7

Khairani, R. N., Cahyadi, S., & Susanto, H. (2021). Studi validitas dan reliabilitas mindfulness in teaching scale Indonesia [Validity and reliability study of the mindfulness in teaching scale Indonesia]. *Jurnal Kependidikan: Jurnal Hasil Penelitian Dan Kajian Kepustakaan Di Bidang Pendidikan, Pengajaran Dan Pembelajaran*, *7*(1), 55–63. https://doi.org/10.33394/jk.v7i1.3296

Koh, J. H. L. (2013). A rubric for assessing teachers' lesson activities with respect to TPACK for meaningful learning with ICT. *Australasian Journal of Educational Technology*, *29*(6), 887–900. https://doi.org/10.14742/ajet.228

Kovač, V. B., Nome, D. Ø., Jensen, A. R., & Skreland, L. L. (2023). The why, what and how of deep learning: Critical analysis and additional concerns. *Education Inquiry*, *16*(2), 237–253. https://doi.org/10.1080/20004508.2023.2194502

Lee, C. B. (2018). Initial development of the meaningful learning with technology scale (MeLTS) for high-school students. *Interactive Learning Environments*, *26*(2), 163–174. https://doi.org/10.1080/10494820.2017.1283336

Leppma, M., & Darrah, M. (2024). Self-efficacy, mindfulness, and self-compassion as predictors of math anxiety in undergraduate students. *International Journal of Mathematical Education in Science and Technology*, *55*(4), 883–898. https://doi.org/10.1080/0020739x.2022.2054740

Ling Lee, W., Chinna, K., & Sumintono, B. (2021). Psychometrics assessment of HeartQoL questionnaire: A Rasch analysis. *European Journal of Preventive Cardiology*, *28*(12), e1–e5. https://doi.org/10.1177/2047487320902322

Medvedev, O. N., & Krägeloh, C. U. (2022). Rasch measurement model. In O. N. Medvedev, C. U. Krägeloh, R. J. Siegert, & N. N. Singh (Eds.), *Handbook of Assessment in Mindfulness Research* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-030-77644-2_4-1

Meika, I., Sartika, N. S., Sujana, A., Jarinah, J., Hakim, Z., Windiarti, I. S., & Hendra, H. (2025). E-didactics design of differential calculus based on TPACK to overcome learning obstacles for mathematics pre-service teachers. *Infinity Journal*, *14*(3), 733–752. https://doi.org/10.22460/infinity.v14i3.p733-752

Mooduto, Y. S., & Musa, A. V. (2021). Improving student learning outcomes using the class IV joyful learning method in elementary school. *Journal of Asian Multicultural Research for Educational Study*, *2*(1), 29–35.

Mudrikah, M., Agustiningsih, S., Shubah, L., & Richardo, R. (2025). Analysis of joyful learning to a growth mindset in learning mathematics. In *Proceedings of the 1st Alma Ata International Conference on Education (AAICE 2023)*, (pp. 23–33). https://doi.org/10.2991/978-2-38476-414-3_3

Muhtarom, M. (2024). Developing an instruments to measure prospective teacher beliefs about mathematical problem-solving using the Rasch model. *Jurnal Elemen*, *10*(2), 274–288. https://doi.org/10.29408/jel.v10i2.25040

Muhtarom, M., Siswono, T. Y. E., & Juniati, D. (2020). Climber prospective teacher: Relationship beliefs and mathematics teaching practice. *Al-Jabar : Jurnal Pendidikan Matematika*, *11*(1), 83–92. https://doi.org/10.24042/ajpm.v11i1.6103

Nafi'ah, J., & Faruq, D. J. (2025). Conceptualizing deep learning approach in primary education: Integrating mindful, meaningful, and joyful. *Journal of Educational Research and Practice*, *3*(2), 225–237. https://doi.org/10.70376/jerp.v3i2.384

Nieveen, N., & Folmer, E. (2013). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), *Educational design research - Part A: An introduction* (pp. 152–169). Netherlands Institute for Curriculum Development (SLO).

O'Neill, T. R., Gregg, J. L., & Peabody, M. R. (2020). Effect of sample size on common item equating using the dichotomous Rasch model. *Applied Measurement in Education*, *33*(1), 10–23. https://doi.org/10.1080/08957347.2019.1674309

Ojell, H., Palohuhta, M., & Ferreira, J. M. (2023). A qualitative microanalysis of the immediate behavioural effects of mindfulness practices on students' self-regulation and attention. *Trends in Psychology*, *31*(4), 641–664. https://doi.org/10.1007/s43076-023-00263-1

Othman, H., Asshaari, I., Bahaludin, H., Nopiah, Z. M., & Ismail, N. A. (2012). Application of Rasch measurement model in reliability and quality evaluation of examination paper for engineering mathematics courses. *Procedia - Social and Behavioral Sciences*, *60*, 163–171. https://doi.org/10.1016/j.sbspro.2012.09.363

Perrotta, C., & Selwyn, N. (2020). Deep learning goes to school: Toward a relational understanding of AI in education. *Learning, Media and Technology*, *45*(3), 251–269. https://doi.org/10.1080/17439884.2020.1686017

Purnomo, Y. W. (2017). A scale for measuring teachers' mathematics-related beliefs: A validity and reliability study. *International Journal of Instruction*, *10*(2), 23–38.

Raihan, M. A., & Han, S. L. (2012). Technology integration for meaningful learning-the constructivist view. *Bangladesh Educational Journal*, *11*(1), 17–37.

Rezaei, A. (2011). Relationship between epistemological beliefs and conceptions of learning in students with surface-and deep-learning approaches. *Advances in Cognitive Science*, *13*(1), 1–14.

Roeser, R. W., Skinner, E., Beers, J., & Jennings, P. A. (2012). Mindfulness training and teachers' professional development: An emerging area of research and practice. *Child Development Perspectives*, *6*(2), 167–173. https://doi.org/10.1111/j.1750-8606.2012.00238.x

Setiawan, B., Panduwangi, M., & Sumintono, B. (2018). A Rasch analysis of the community's preference for different attributes of Islamic banks in Indonesia. *International Journal of Social Economics*, *45*(12), 1647–1662. https://doi.org/10.1108/ijse-07-2017-0294

Sinha, A., & Ayush, K. (2018). Towards mathematical reasoning: A multimodal deep learning approach. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, (pp. 4028–4032). https://doi.org/10.1109/icip.2018.8451353

Siswanto, D. H., & Afandi, M. M. (2024). Analysis of the implementation of mindfulness in senior high school students' learning. *Asian Pendidikan*, *4*(1), 79–85. https://doi.org/10.53797/aspen.v4i1.9.2024

Siswono, T. Y. E., Kohar, A. W., & Hartono, S. (2019). Beliefs, knowledge, teaching practice: three factors affecting the quality of teacher's mathematical problem-solving. *Journal of Physics: Conference Series*, *1317*(1), 012127. https://doi.org/10.1088/1742-6596/1317/1/012127

Siswono, T. Y. E., Kohar, A. W., Hartono, S., Rosyidi, A. H., Kurniasari, I., & Karim, K. (2019). Examining teacher mathematics-related beliefs and problem-solving knowledge for teaching: Evidence from Indonesian primary and secondary teachers. *lnternational Electronic Journal of Elementary Education*, *11*(5), 493–506. https://doi.org/10.26822/iejee.2019553346

Sulistyowati, F., Istiqomah, I., Putri, N. A. H., Septiani, D., Kusumaningrum, B., & Kuncoro, K. S. (2024). Interactive e-learning using web-wix for joyful learning linear equations. *AIP Conference Proceedings*, *3046*(1), 020033. https://doi.org/10.1063/5.0194657

Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. In *Proceedings of the 1st International Conference on Education Innovation (ICEI 2017)*, (pp. 38–42). https://doi.org/10.2991/icei-17.2018.11

Ulfa, N., & Hidayat, A. (2024). Development of mathematics teaching modules based on joyfull learning. In *Proceedings Series of Educational Studies*, (pp. 9–15).

Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*, *32*(1), 108–114. https://doi.org/10.7334/psicothema2019.286

Wei Chan, S., Ismail, Z., & Sumintono, B. (2015). The impact of statistical reasoning learning environment: A Rasch analysis. *Advanced Science Letters*, *21*(5), 1211–1215. https://doi.org/10.1166/asl.2015.6077

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334. https://doi.org/10.1037/a0032121

Appendix

"*Beliefs about the implementation of deep learning*" scale for mathematics teachers

Note: (+) is favorable item, (-) is unfavorable item

1. I understand my emotional state when teaching mathematics by paying attention to lesson planning. (+)
2. I prioritize delivering mathematics material over addressing students' emotional states. (-)
3. I implement mathematics learning activities that encourage collaboration between students and teachers in understanding mathematical concepts. (+)
4. I implement mathematics learning activities that make students passive recipients of the material. (-)
5. I present mathematics material abstractly without connecting it to students' everyday experiences. (-)
6. I ignore personal feelings when facing difficult situations in the mathematics teaching process. (-)
7. I listen carefully to students' ideas, even when I disagree with them. (+)
8. I implement mathematics learning by focusing the learning process on the teacher's explanations and the teacher as the sole source of information. (-)
9. I use learning strategies such as discussions, games, or math projects, especially for students who quickly grasp the material, to make learning activities more effective. (+)
10. I present real-life problems from everyday life as part of mathematics learning in class. (+)
11. I presented mathematics material without linking it to concepts students had previously mastered. (-)
12. I guided students to design their own learning goals and strategies for understanding mathematical concepts. (+)
13. I structured the mathematics learning process with a more formal and structured approach in class. (-)
14. I designed mathematics learning activities that could arouse students' curiosity and enthusiasm for learning. (+)
15. I conducted mathematics lessons with a primary focus on delivering the material without any particular emphasis on recognizing student contributions or progress. (-)
16. I helped students develop an understanding of mathematical concepts by integrating their existing knowledge. (+)
17. I directed learning without giving students space to determine their own learning goals. (-)
18. I created a mathematics learning environment that made students feel comfortable, enthusiastic, and actively participated. (+)
19. I presented mathematics material in a style that tended to be monotonous for some students. (-)
20. I appreciated students' efforts and achievements in understanding mathematics material. (+)