Manuscript Title : Quality of Four-Tier Diagnostic Test on Wave and Vibration Materials: An Empirical Study Using Rasch Modeling

Author : Moh. Irma Sukarelawan, **Toni Kus Indratno**, Raden Oktova, Nurul Syafiqah Yap Abdullah

Journal : Jurnal Pendidikan Fisika Indonesia

1. Submitted to the journal "Jurnal Pendidikan Fisika Indonesia" – 08 September 2023
2. Revisions required, initial review from Editor – 12 September 2023
3. Revised version submitted – 14 September 2023
4. First revisions, review from Reviewer A – 04 December 2023
5. First revised version submitted – 04 January 2024
6. Second revisions, review from Reviewer A – 16 January 2024
7. Second revised version submitted – 26 January 2024
8. Third revisions, review from Reviewer A – 05 February 2024
9. Third revisions, review from Reviewer B – 15 February 2024
10. Third revised version submitted – 08 March 2024
11. Forth revisions, review from Reviewer B – 25 March 2024
12. Forth revised version submitted – 27 March 2024
13. Paper Accepted – 23 April 2024
14. Paper published – 01 July 2024

1. Submitted to the journal “Jurnal Pendidikan Fisika Indonesia”

08 September 2023

UNIVERSITAS
AHMAD DAHLAN

Toni Kus Indratno <tonikus@staff.uad.ac.id>

[JPFI] Submission Acknowledgement

1 pesan

Prof. Dr. Wiyanto <jpfi@mail.unnes.ac.id> 8 September 2023 pukul 13.25
Kepada: Toni Kus Indratno <tonikus@staff.uad.ac.id>

Toni Kus Indratno:

Thank you for submitting the manuscript, "ARE 4WADI BE APPLIED TO HIGH SCHOOL STUDENTS IN THE INDONESIAN CONTEXT? EMPIRICAL STUDIES IN RASCH MODELING PERSPECTIVE" to Jurnal Pendidikan Fisika Indonesia. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Manuscript URL:
https://journal.unnes.ac.id/nju/index.php/JPFI/author/submission/47493
Username: tonikus

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Prof. Dr. Wiyanto
Jurnal Pendidikan Fisika Indonesia

Jurnal Pendidikan Fisika Indonesia
http://journal.unnes.ac.id/index.php/JPFI

2. Revisions required, initial review from Editor

12 September 2023

UNIVERSITAS
AHMAD DAHLAN

Toni Kus Indratno <tonikus@staff.uad.ac.id>

[JPFI] INFO

1 pesan

Dra. Siti Khanafiyah <sitikhanafiyah@journal.unnes.ac.id> 12 September 2023 pukul 12.01
Kepada: Toni Kus Indratno <tonikus@staff.uad.ac.id>
Cc: "Moh. Irma Sukarelawan" <irma.sukarelawan@pfis.uad.ac.id>

Dear authors (ID 47493),

Please check the JPFI OJS system to get the latest manuscript review results
and upload your corrections immediately.

https://journal.unnes.ac.id/nju/index.php/JPFI

Regards
JPFI

Jurnal Pendidikan Fisika Indonesia
http://journal.unnes.ac.id/index.php/JPFI

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

ARE 4WADI BE APPLIED TO HIGH SCHOOL STUDENTS IN THE INDONESIAN CONTEXT? EMPIRICAL STUDIES IN RASCH MODELING PERSPECTIVE

Toni Kus Indratno[1]*, Moh. Irma Sukarelawan[2], Nurul Syafiqah Yap Abdullah[3]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[3]Department of Physics
Universiti Pendidikan Sultan Idris (UPSI), Malaysia
syafiqah@fsmt.upsi.edu.my

Commented [A1]: Perhatikan kembali tatacara penulisan nama author sesuaikan dengan panduan manuskrip JPFI. Misalkan Toni Kus Indratno menjadi T.K.Indratno

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop a misconception diagnostic instrument. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception diagnostic instruments to vibration and wave materials. The adaptation process was conducted on 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test, I-4WADI was tested for readability on 35 students. Empirical validation data were analyzed using Rasch modeling. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia.

Commented [A2]: Tambagkan manfaat penelitian

Key words: contains 3 to 5 main words and should be listed alphabetically

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Aykutlu, Ensari, & Bayrak, 2023; Bhagat, Subheesh, Bhattacharya, & Chang, 2017; BURGE, 1967; Çelikkanlı & Kızılcık, 2022; Erickson, 1979; Lemmer, Kriek, & Erasmus, 2020; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Temiz & Yavuz, 2014; Tsui & Treagust, 2010; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Asikainen & Hirvonen, 2009; Bani-Salameh, 2017; Brown, 1992; BURGE, 1967; Chu, Treagust, Yeo, & Zadnik, 2012; Erickson, 1979; Hewson, 1985; Kaniawati et al., 2019; Yeo & Zadnik, 2001). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Caleon & Subramaniam, 2010a, 2010b, 2010c; Kaya Şengören, Tanel, & Kavcar, 2009; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Lim & Poo, 2021; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020; Marista et al., 2023), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials (4WADI) was previously carried out by Caleon and Subramaniam (2010c). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Guo, Lu, Johnson, & McCaffrey, 2022; Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

This study's main instrument was a four-level misconception diagnostic instrument on vibration and wave material consisting of 12 items. The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Puspitasari, Rahmatika, et al., 2022; Sukarelawan, Jumadi, et al., 2021). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULT AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Beaton et al. (2000). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity (Huddleston, 1956). Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Aiken, 1985). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010c).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010c).

3. Empirical test results

a. Item reliability

Commented [A3]: Carilah referensi terbaru, maksimal referensi 10 tahun yang lalu, jika melebihi rentang tersebut maka perlu dicari jurnal terbaru yang relevan dengan penelitian saudara

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Wahab & Tentama, 2020; Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (Caleon & Subramaniam, 2010c).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVP and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is room for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item Difficulty Distribution (LVP and Wright map)

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Rusmansyah & Almubarak, 2020; Sumintono & Widhiarso, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

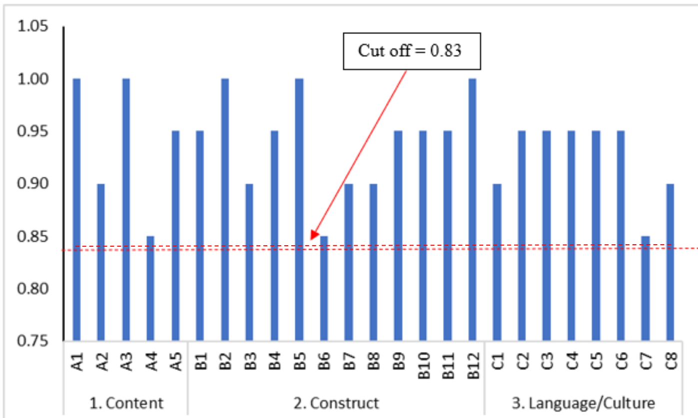

Figure 1 Aiken V validity coefficient

Commented [A4]: Perjelas kembali keterangan untuk x dan y

```
    MEASURE    Person - MAP - Item
                   <more>|<rare>
        1             .  +T
                      .  |
                      #  |
                      .  |  Q11
                      #  |  Q1
                      .  |S Q5
                      .  |  Q4
                      . T|
                         |  Q8
                      #  |
                         |M Q2      Q7
                    ####  |  Q10     Q6
        0            .## S+  Q3
                   .####  |
                    .###  |
                   .####  |S Q9
                  ######  |
                 .###### M|
           #############  |
               .########  |
                 #######  |T
                         |
                    .### S|
                     .##  |  Q12
                         |
       -1           ####  +
                         |
                        T|
                      .  |
                         |
                         |
                      .  |
                         |
                         |
                         |
```

```
                              |
                              |
      -2                 .  +
                    <less>|<freq>
     EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 1. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	16.7485	100.0%		100.0%
Raw variance explained by measures	=	4.7485	28.4%		28.4%
Raw variance explained by persons	=	.5610	3.3%		3.4%
Raw Variance explained by items	=	4.1875	25.0%		25.1%
Raw unexplained variance (total)	=	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast	=	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast	=	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast	=	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast	=	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast	=	1.1818	7.1%	9.8%	

d. The unidimensionality of the items

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument

measures one single dimension in the latent factors investigated in this study (Hagquist, Bruce, & Gustavsson, 2009; Kreiner, 2007; Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. Retrieved from https://doi.org/10.1177/0013164485451012

Asikainen, M. A., & Hirvonen, P. E. (2009). A Study of Pre- and Inservice Physics Teachers ' Understanding of Photoelectric Phenomenon as Part of the Development of a Research-Based Quantum Physics Course. *American Journal of Physics*, 77(7), 658–666.

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, 25(24), 3186–3191. Retrieved from https://doi.org/10.1097/00007632-200012150-00014

Bhagat, K. K., Subheesh, N. P., Bhattacharya, B., & Chang, C.-Y. (2017). The Design and Development of Identification of Students' Misconceptions in Individualized Learning Environment (iSMILE)

Commented [A5]: Referensi yang dapat digunakan rentang 10 tahun yang lalu, jika melebihi rentang tersebut sebaiknya carilah jurnal terbaru yang relefan dengan penelitian anda

System. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(1), 19–34. Retrieved from https://doi.org/10.12973/eurasia.2017.00602a

Brown, D. E. (1992). Using Examples and Analogies to Remediate Misconceptions in Physics: Factors Influencing .Conceptual Change. *Journal of Research in Science Teaching*, 29(1), 17–34.

BURGE, E. J. (1967). Misconceptions in nuclear physics. *Physics Education*, 2, 184–187.

Caleon, I. S., & Subramaniam, R. (2010a). Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves. *International Journal of Science Education*, 32(7), 939–961. Retrieved from https://doi.org/10.1080/09500690902890130

Caleon, I. S., & Subramaniam, R. (2010b). Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves. *International Journal of Science Education*, 32(7), 939–961. Retrieved from https://doi.org/10.1080/09500690902890130

Caleon, I. S., & Subramaniam, R. (2010c). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Chu, H.-E., Treagust, D. F., Yeo, S., & Zadnik, M. (2012). Evaluation of Students ' Understanding of Thermal Concepts in Everyday Contexts. *International Journal of Science Education*, 34(10), 1509–1534. Retrieved from https://doi.org/10.1080/09500693.2012.657714

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Erickson, G. L. (1979). Children ' s Conceptions of Heat and Temperature. *Science Education*, 63(2), 221–230.

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, 46(3), 380–393. Retrieved from https://doi.org/10.1016/j.ijnurstu.2008.10.007

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

Hewson, P. W. (1985). Diagnosis and remediation of an alternative conception of velocity using a microcomputer program. *American Journal of Physics*, 53(7), 684–690. Retrieved from https://doi.org/10.1119/1.14285

Huddleston, E. M. (1956). Test Development on the Basis of Content Validity. *Educational and Psychological Measurement*, 16(3), 283–293. Retrieved from https://doi.org/10.1177/001316445601600302

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. Retrieved from https://doi.org/10.12973/tused.10269a

Kaya Şengören, S., Tanel, R., & Kavcar, N. (2009). Students difficulties about the wave pulses propagating on a rope. *Journal of Turkish Science Education*, 6(1), 50–59.

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, 59(3), 268–298. Retrieved from https://doi.org/10.1027/1901-2276.59.3.268

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Marista, J., Latifah, S., Saregar, A., Anwar, C., Jamaluddin, W., & Fitri, M. R. (2023). Physics E-poster to reduce the risk of misconceptions on temperature and heat: Students' perception (p. 020001). Retrieved from https://doi.org/10.1063/5.0124209

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian

Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Temiz, B. K., & Yavuz, A. (2014). Students' Misconceptions about Newton's Second Law in Outer Space. *European Journal of Physics*, 35(4), 1–16.

Tsui, C. Y., & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, 32(8), 1073–1098. Retrieved from https://doi.org/10.1080/09500690902951429

Wahab, F. I. A., & Tentama, F. (2020). Construct validity and reliability test on burnout. *International Journal of Scientific and Technology Research*, 9(1), 359–363.

Yeo, S., & Zadnik, M. (2001). Introductory thermal concept evaluation: assessing students' understanding. *The Physics Teacher*, 39(8), 496–504. Retrieved from https://doi.org/10.1119/1.1424603

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

3. Revised version submitted

14 September 2023

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

ARE 4WADI BE APPLIED TO HIGH SCHOOL STUDENTS IN THE INDONESIAN CONTEXT? EMPIRICAL STUDIES IN RASCH MODELING PERSPECTIVE

T. K Indratno[1]*, M. I. Sukarelawan[2]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop a misconception diagnostic instrument. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception diagnostic instruments to vibration and wave materials. The adaptation process was conducted on 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test, I-4WADI was tested for readability on 35 students. Empirical validation data were analyzed using Rasch modeling. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who want to diagnose student misconceptions regarding vibrations and waves.

Key words: contains 3 to 5 main words and should be listed alphabetically

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Aykutlu et al., 2023; Bhagat et al., 2017; BURGE, 1967; Çelikkanlı & Kızılcık, 2022; Erickson, 1979; Lemmer et al., 2020; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Temiz & Yavuz, 2014; Tsui & Treagust, 2010; Zhao et al., 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Assimi et al., 2022; Bani-Salameh, 2017; Kaniawati et al., 2019; Kusairi et al., 2022; Rismaningsih & Nurhafsari, 2022; Suhendi & Ardiansyah, 2021; Sukarelawan, Puspitasari, Sulisworo, et al., 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Caleon & Subramaniam, 2010a, 2010b; Kaya Şengören et al., 2009; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Assimi et al., 2022; Lim & Poo, 2021; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar et al., 2020; Marista et al., 2023), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial et al., 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, et al., 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials (4WADI) was previously carried out by Caleon and Subramaniam (2010b). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Guo et al., 2022; Sukarelawan, Jumadi, et al., 2021). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee et al., 2020).

This study's main instrument was a four-level misconception diagnostic instrument on vibration and wave material consisting of 12 items. The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna et al., 2021). Then, use the Logit Value of Item combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULT AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz et al. (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity (Huddleston, 1956). Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00.

This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Aiken, 1985). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010b).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010b).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Wahab & Tentama, 2020; Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (Caleon & Subramaniam, 2010b).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams et al., 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVP and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item Difficulty Distribution (LVP and Wright map)

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Rusmansyah & Almubarak, 2020; Sumintono & Widhiarso, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

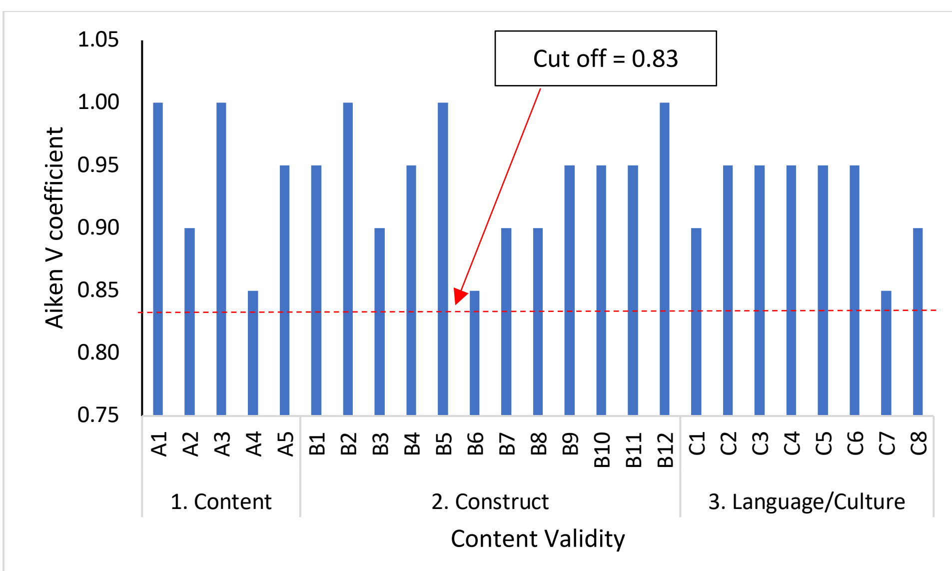

Figure 1. Aiken V validity coefficient

```
MEASURE    Person - MAP - Item
               <more>|<rare>
    1             .  +T
                  .  |
                  #  |
                  .  |  Q11
                  #  |  Q1
                  .  |S Q5
                  .  |  Q4
                  . T|
                     |  Q8
                  #  |
                     |M Q2     Q7
               ####  |  Q10    Q6
    0           .## S+  Q3
              .####  |
               .###  |
              .####  |S Q9
             ######  |
            .###### M|
      #############  |
          .########  |
            #######  |T
                     |
               .### S|
                .##  |  Q12
                     |
   -1          ####  +
                     |
                    T|
                  .  |
                     |
                     |
                  .  |
                     |
                     |
                     |
                     |
                     |
   -2             .  +
               <less>|<freq>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	16.7485	100.0%		100.0%
Raw variance explained by measures	=	4.7485	28.4%		28.4%
Raw variance explained by persons	=	.5610	3.3%		3.4%
Raw Variance explained by items	=	4.1875	25.0%		25.1%
Raw unexplained variance (total)	=	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast	=	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast	=	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast	=	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast	=	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast	=	1.1818	7.1%	9.8%	

d. The unidimensionality of the items

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Hagquist et al., 2009; Kreiner, 2007; Scoulas et al., 2021; Simpelaere et al., 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, *18*(4), 515–534. https://doi.org/10.1108/ITSE-03-2020-0032

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, *44*(1), 015701. https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, *52*(1), 14003.

Bhagat, K. K., Subheesh, N. P., Bhattacharya, B., & Chang, C.-Y. (2017). The Design and Development of Identification of Students' Misconceptions in Individualized Learning Environment (iSMILE) System. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(1), 19–34. https://doi.org/10.12973/eurasia.2017.00602a

BURGE, E. J. (1967). Misconceptions in nuclear physics. *Physics Education*, *2*, 184–187.

Caleon, I. S., & Subramaniam, R. (2010a). Exploring Students' Conceptualization of the Propagation of Periodic Waves. *The Physics Teacher*, *48*(55), 55–59. https://doi.org/10.1119/1.3274366

Caleon, I. S., & Subramaniam, R. (2010b). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485–493. https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., Karaduman, A., Swan, K., Clavé, P., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, *33*(4), 441–456. https://doi.org/10.1007/s00455-017-9873-4

Erickson, G. L. (1979). Children ' s Conceptions of Heat and Temperature. *Science Education*, *63*(2), 221–230.

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, *2022*(1), 1–16. https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), 16. https://doi.org/10.1186/s40468-021-00132-7

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, *46*(3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 540–553. https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, *17*(9), e0271992. https://doi.org/10.1371/journal.pone.0271992

Huddleston, E. M. (1956). Test Development on the Basis of Content Validity. *Educational and Psychological Measurement*, *16*(3), 283–293. https://doi.org/10.1177/001316445601600302

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

Kaya Şengören, S., Tanel, R., & Kavcar, N. (2009). Students difficulties about the wave pulses propagating on a rope. *Journal of Turkish Science Education*, *6*(1), 50–59.

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, *59*(3), 268–298. https://doi.org/10.1027/1901-2276.59.3.268

Kusairi, S., Sutopo, .., & Suryadi, A. (2022). Utilizing Isomorphic Multiple-Choice Items to Diagnose Students' Misconceptions in Force and Motion. *New Physics: Sae Mulli*, *72*(4), 311–318. https://doi.org/10.3938/NPSM.72.311

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, *50*(2), 375–392. https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., Zhang, W., & Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, *in press*. https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, *10*(2), 241–252. https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, *11*(August), 1–18. https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, *14*(2). https://doi.org/10.30880/jtet.2022.14.02.005

Marista, J., Latifah, S., Saregar, A., Anwar, C., Jamaluddin, W., & Fitri, M. R. (2023). *Physics E-poster to reduce the risk of misconceptions on temperature and heat: Students' perception*. 020001. https://doi.org/10.1063/5.0124209

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, *41*(2), 229–236. https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., Ceroni, L., Sonnery-Cottet, B., & Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, *23*(1), 1–8. https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *5*(2), 69–76. https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, *in press*. https://doi.org/10.47626/2237-6089-2023-0630

Rismaningsih, F., & Nurhafsari, A. (2022). Identify Hydrostatic Misconceptions Using Four Tier Diagnostic Tests With the Help of iSpring Suite 9. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *5*(3), 222. https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, *43*(1), 101071. https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, *15*(1), 66. https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, *8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, *1918*(2). https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, *10*(4), 486–495. https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., Indratno, T. K., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, *10*(2), 216–224. https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, *55*(1), 127–141. https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, *5*(1), 1–8. https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*.

Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, *88*(2), 211–221. https://doi.org/10.2298/JSC220117073S

Temiz, B. K., & Yavuz, A. (2014). Students' Misconceptions about Newton's Second Law in Outer Space. *European Journal of Physics*, *35*(4), 1–16.

Tsui, C. Y., & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, *32*(8), 1073–1098. https://doi.org/10.1080/09500690902951429

Wahab, F. I. A., & Tentama, F. (2020). Construct validity and reliability test on burnout. *International Journal of Scientific and Technology Research*, *9*(1), 359–363.

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 35–47. https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, *57*(1), 51–67. https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, *17*(1), 1–12. https://doi.org/10.1186/s12955-019-1168-z

4. First revisions, review from Reviewer A

04 December 2023

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

ARE 4WADI BE APPLIED TO HIGH SCHOOL STUDENTS IN THE INDONESIAN CONTEXT? EMPIRICAL STUDIES IN RASCH MODELING PERSPECTIVE

Commented [A1]: The title in the statement sentence is not a question

T. K Indratno[1]*, M. I. Sukarelawan[2]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Commented [A2]: Added the method

Not all teachers have the ability to develop a misconception diagnostic instrument. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception diagnostic instruments to vibration and wave materials. The adaptation process was conducted on 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test, I-4WADI was tested for readability on 35 students. Empirical validation data were analyzed using Rasch modeling. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who want to diagnose student misconceptions regarding vibrations and waves.

Key words: contains 3 to 5 main words and should be listed alphabetically

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Aykutlu et al., 2023; Bhagat et al., 2017; BURGE, 1967; Çelikkanlı & Kızılcık, 2022; Erickson, 1979; Lemmer et al., 2020; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Temiz & Yavuz, 2014; Tsui & Treagust, 2010; Zhao et al., 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Assimi et al., 2022; Bani-Salameh, 2017; Kaniawati et al., 2019; Kusairi et al., 2022; Rismaningsih & Nurhafsari, 2022; Suhendi & Ardiansyah, 2021; Sukarelawan, Puspitasari, Sulisworo, et al., 2022). When students develop a wrong understanding of key concepts in physics,

Commented [A3]: The last ten years

this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Caleon & Subramaniam, 2010a, 2010b; Kaya Şengören et al., 2009; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Assimi et al., 2022; Lim & Poo, 2021; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisw
oro, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar et al., 2020; Marista et al., 2023), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial et al., 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, et al., 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials (4WADI) was previously carried out by Caleon and Subramaniam (2010b). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Guo et al., 2022; Sukarelawan, Jumadi, et al., 2021). So, another

Commented [A4]: The last ten years

Commented [A6]: adjust the journal's surrounding style

Commented [A7]: adjust the journal's surrounding style

Commented [A5]: adjust the journal's surrounding style

Commented [A8]: adjust the journal's surrounding style approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee et al., 2020).

This study's main instrument was a four-level misconception diagnostic instrument on vibration and wave material consisting of 12 items. The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna et al., 2021). Then, use the Logit Value of Item combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

Commented [A9]: Add a taxonomy for each question

Commented [A10]: adjust the journal's surrounding style

RESULT AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz et al. (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity (Huddleston, 1956). Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00.

This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Aiken, 1985). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010b).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010b).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Wahab & Tentama, 2020; Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (Caleon & Subramaniam, 2010b).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams et al., 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVP and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item Difficulty Distribution (LVP and Wright map)

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Rusmansyah & Almubarak, 2020; Sumintono & Widhiarso, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

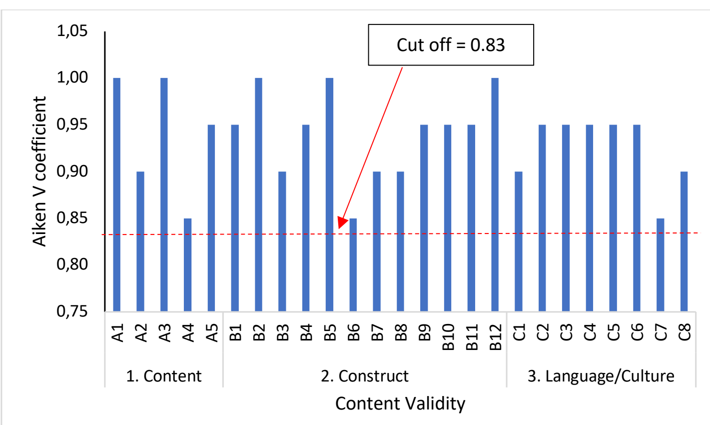

Figure 1. Aiken V validity coefficient

```
    MEASURE    Person - MAP - Item
                   <more>|<rare>
      1              .  +T
                     .  |
                     #  |
                     .  |  Q11
                     #  |  Q1
                     .  |S Q5
                     .  |  Q4
                     . T|
                        |  Q8
                     #  |
                        |M Q2     Q7
                  ####  |  Q10    Q6
      0            .## S+  Q3
                 .####  |
                  .###  |
                 .####  |S Q9
                ######  |
               .###### M|
         #############  |
             .########  |
               #######  |T
                        |
                  .### S|
                   .##  |  Q12
                        |
     -1           ####  +
                        |
                       T|
                     .  |
                        |
                        |
                     .  |
                        |
                        |
                        |
                        |
                        |
     -2              .  +
                   <less>|<freq>
     EACH "#" IS 4: EACH "." IS 1 TO 3
```
Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

d. The unidimensionality of the items

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Hagquist et al., 2009; Kreiner, 2007; Scoulas et al., 2021; Simpelaere et al., 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it

Commented [A11]: Times new roman

is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, *18*(4), 515–534. https://doi.org/10.1108/ITSE-03-2020-0032

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, *44*(1), 015701. https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, *52*(1), 14003.

Bhagat, K. K., Subheesh, N. P., Bhattacharya, B., & Chang, C.-Y. (2017). The Design and Development of Identification of Students' Misconceptions in Individualized Learning Environment (iSMILE) System. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(1), 19–34. https://doi.org/10.12973/eurasia.2017.00602a

BURGE, E. J. (1967). Misconceptions in nuclear physics. *Physics Education*, *2*, 184–187.

Caleon, I. S., & Subramaniam, R. (2010a). Exploring Students' Conceptualization of the Propagation of Periodic Waves. *The Physics Teacher*, *48*(55), 55–59. https://doi.org/10.1119/1.3274366

Caleon, I. S., & Subramaniam, R. (2010b). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An

Commented [A12]: 12 letter font and last 10 years , The font size and type are made the same

example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485–493. https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., Karaduman, A., Swan, K., Clavé, P., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, *33*(4), 441–456. https://doi.org/10.1007/s00455-017-9873-4

Erickson, G. L. (1979). Children ' s Conceptions of Heat and Temperature. *Science Education*, *63*(2), 221–230.

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, *2022*(1), 1–16. https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), 16. https://doi.org/10.1186/s40468-021-00132-7

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, *46*(3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 540–553. https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, *17*(9), e0271992. https://doi.org/10.1371/journal.pone.0271992

Huddleston, E. M. (1956). Test Development on the Basis of Content Validity. *Educational and Psychological Measurement*, *16*(3), 283–293. https://doi.org/10.1177/001316445601600302

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

Kaya Şengören, S., Tanel, R., & Kavcar, N. (2009). Students difficulties about the wave pulses propagating on a rope. *Journal of Turkish Science Education*, *6*(1), 50–59.

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, *59*(3), 268–298. https://doi.org/10.1027/1901-2276.59.3.268

Kusairi, S., Sutopo, .., & Suryadi, A. (2022). Utilizing Isomorphic Multiple-Choice Items to Diagnose Students' Misconceptions in Force and Motion. *New Physics: Sae Mulli*, *72*(4), 311–318. https://doi.org/10.3938/NPSM.72.311

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, *50*(2), 375–392. https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., Zhang, W., & Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, *in press*. https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, *10*(2), 241–252. https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, *11*(August), 1–18. https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, *14*(2). https://doi.org/10.30880/jtet.2022.14.02.005

Marista, J., Latifah, S., Saregar, A., Anwar, C., Jamaluddin, W., & Fitri, M. R. (2023). *Physics E-poster to reduce the risk of misconceptions on temperature and heat: Students' perception*. 020001. https://doi.org/10.1063/5.0124209

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, *41*(2), 229–236. https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., Ceroni, L., Sonnery-Cottet, B., & Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, *23*(1), 1–8. https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *5*(2), 69–76. https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, *in press*. https://doi.org/10.47626/2237-6089-2023-0630

Rismaningsih, F., & Nurhafsari, A. (2022). Identify Hydrostatic Misconceptions Using Four Tier Diagnostic Tests With the Help of iSpring Suite 9. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *5*(3), 222. https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, *43*(1), 101071. https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, *15*(1), 66. https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, *8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, *1918*(2). https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, *10*(4), 486–495. https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., Indratno, T. K., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, *10*(2), 216–224. https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, *55*(1), 127–141. https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, *5*(1), 1–8. https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*.

Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, *88*(2), 211–221. https://doi.org/10.2298/JSC220117073S

Temiz, B. K., & Yavuz, A. (2014). Students' Misconceptions about Newton's Second Law in Outer Space. *European Journal of Physics*, *35*(4), 1–16.

Tsui, C. Y., & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, *32*(8), 1073–1098. https://doi.org/10.1080/09500690902951429

Wahab, F. I. A., & Tentama, F. (2020). Construct validity and reliability test on burnout. *International Journal of Scientific and Technology Research*, *9*(1), 359–363.

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 35–47. https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, *57*(1), 51–67. https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, *17*(1), 1–12. https://doi.org/10.1186/s12955-019-1168-z

5. First revised version submitted

04 January 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

T. K Indratno[1]*, M. I. Sukarelawan[2]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop misconception-diagnostic instruments. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to vibration and wave materials. The adaptation process was conducted with 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. The expert validation results were analyzed using the Aiken V technique. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about vibrations and waves.

Key words: contains 3 to 5 main words and should be listed alphabetically

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Aykutlu et al., 2023; Bhagat et al., 2017; Brundage & Singh, 2023; Çelikkanlı & Kızılcık, 2022; Lemmer et al., 2020; Stefanou et al., 2023; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Temiz & Yavuz, 2014; Zhao et al., 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Assimi et al., 2022; Bani-Salameh, 2017; Kaniawati et al., 2019; Kusairi et al., 2022;

Rismaningsih & Nurhafsari, 2022; Suhendi & Ardiansyah, 2021; Sukarelawan, Puspitasari, Sulisworo, et al., 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Caleon & Subramaniam, 2010; Negoro & Karina, 2019; Somroob & Wattanakasiwich, 2017). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Assimi et al., 2022; Lim & Poo, 2021; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar et al., 2020; Marista et al., 2023), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial et al., 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, et al., 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials (4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of

Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Guo et al., 2022; Sukarelawan, Jumadi, et al., 2021). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee et al., 2020).

The main instrument of this research is the four-level misconception diagnostic instrument on vibration and wave material (I-4WADI), which consists of 12 items. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna et al., 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz et al. (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity (Huddleston, 1956). Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Aiken, 1985). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Wahab & Tentama, 2020; Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams et al., 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Rusmansyah & Almubarak, 2020; Sumintono & Widhiarso, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

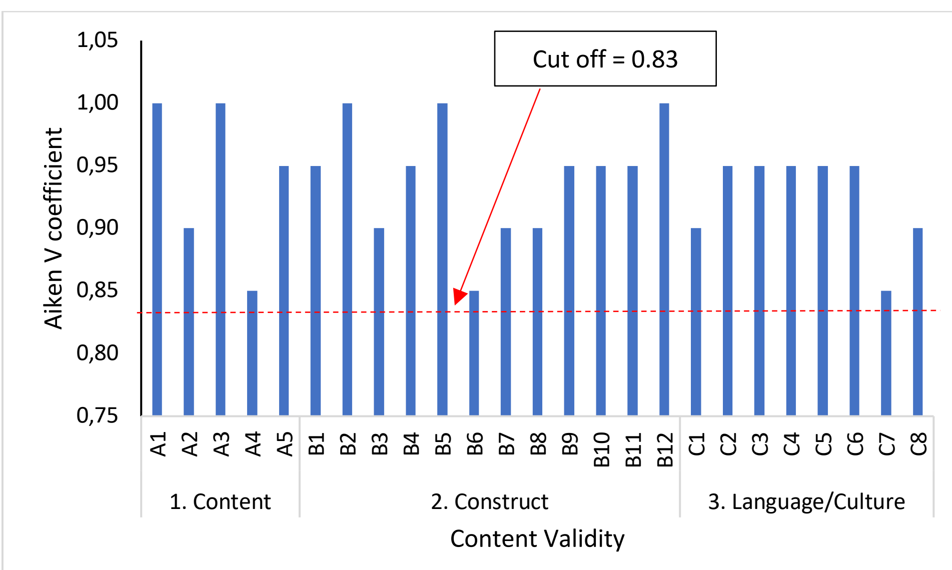

Figure 1. Aiken V validity coefficient

```
   MEASURE    Person - MAP - Item
                   <more>|<rare>
       1               .  +T
                       .  |
                       #  |
                       .  |  Q11
                       #  |  Q1
                       .  |S Q5
                       .  |  Q4
                       . T|
                          |  Q8
                       #  |
                          |M Q2     Q7
                    ####  |  Q10    Q6
       0             .## S+  Q3
                   .####  |
                    .###  |
                   .####  |S Q9
                  ######  |
                 .###### M|
            #############  |
               .########  |
                 #######  |T
                          |
                    .### S|
                     .##  |  Q12
                          |
      -1            ####  +
                          |
                         T|
                       .  |
                          |
                          |
                       .  |
                          |
                          |
                          |
                          |
                          |
      -2               .  +
                   <less>|<freq>
    EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

		Eigenvalue	Observed		Expected
Total raw variance in observations	=	16.7485	100.0%		100.0%
Raw variance explained by measures	=	4.7485	28.4%		28.4%
Raw variance explained by persons	=	.5610	3.3%		3.4%
Raw Variance explained by items	=	4.1875	25.0%		25.1%
Raw unexplained variance (total)	=	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast	=	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast	=	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast	=	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast	=	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast	=	1.1818	7.1%	9.8%	

d. Unidimensionality

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Hagquist et al., 2009; Kreiner, 2007; Scoulas et al., 2021; Simpelaere et al., 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, *18*(4), 515–534. https://doi.org/10.1108/ITSE-03-2020-0032

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, *44*(1), 015701. https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, *52*(1), 14003.

Bhagat, K. K., Subheesh, N. P., Bhattacharya, B., & Chang, C.-Y. (2017). The Design and Development of Identification of Students' Misconceptions in Individualized Learning Environment (iSMILE) System. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(1), 19–34. https://doi.org/10.12973/eurasia.2017.00602a

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, *19*(2), 020112. https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485–493. https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., Karaduman, A., Swan, K., Clavé, P., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, *33*(4), 441–456. https://doi.org/10.1007/s00455-017-9873-4

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, *2022*(1), 1–16. https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), 16. https://doi.org/10.1186/s40468-021-00132-7

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, *46*(3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 540–553. https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, *17*(9), e0271992. https://doi.org/10.1371/journal.pone.0271992

Huddleston, E. M. (1956). Test Development on the Basis of Content Validity. *Educational and Psychological Measurement*, *16*(3), 283–293. https://doi.org/10.1177/001316445601600302

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

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, *59*(3), 268–298. https://doi.org/10.1027/1901-2276.59.3.268

Kusairi, S., Sutopo, .., & Suryadi, A. (2022). Utilizing Isomorphic Multiple-Choice Items to Diagnose Students' Misconceptions in Force and Motion. *New Physics: Sae Mulli*, *72*(4), 311–318. https://doi.org/10.3938/NPSM.72.311

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, *50*(2), 375–392. https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., Zhang, W., & Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, *in press*. https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, *10*(2), 241–252. https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, *11*(August), 1–18. https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, *14*(2). https://doi.org/10.30880/jtet.2022.14.02.005

Marista, J., Latifah, S., Saregar, A., Anwar, C., Jamaluddin, W., & Fitri, M. R. (2023). *Physics E-poster to reduce the risk of misconceptions on temperature and heat: Students' perception*. 020001. https://doi.org/10.1063/5.0124209

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, *41*(2), 229–236. https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., Ceroni, L., Sonnery-Cottet, B., & Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, *23*(1), 1–8. https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *5*(2), 69–76. https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, *in press*. https://doi.org/10.47626/2237-6089-2023-0630

Rismaningsih, F., & Nurhafsari, A. (2022). Identify Hydrostatic Misconceptions Using Four Tier Diagnostic Tests With the Help of iSpring Suite 9. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *5*(3), 222. https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, *43*(1), 101071. https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, *15*(1), 66. https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, *8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, *901*, 012123. https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, *31*(1), 127–157. https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, *1918*(2). https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, *10*(4), 486–495. https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., Indratno, T. K., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, *10*(2), 216–224. https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, *55*(1), 127–141. https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, *5*(1), 1–8. https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, *88*(2), 211–221. https://doi.org/10.2298/JSC220117073S

Temiz, B. K., & Yavuz, A. (2014). Students' Misconceptions about Newton's Second Law in Outer Space. *European Journal of Physics*, *35*(4), 1–16.

Wahab, F. I. A., & Tentama, F. (2020). Construct validity and reliability test on burnout. *International Journal of Scientific and Technology Research*, *9*(1), 359–363.

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 35–47. https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, *57*(1), 51–67. https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, *17*(1), 1–12. https://doi.org/10.1186/s12955-019-1168-z

6. Second revisions, review from Reviewer A

16 January 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

T. K Indratno[1]*, M. I. Sukarelawan[2]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop misconception-diagnostic instruments. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to vibration and wave materials. The adaptation process was conducted with 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. The expert validation results were analyzed using the Aiken V technique. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about vibrations and waves.

Key words: contains 3 to 5 main words and should be listed alphabetically

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Aykutlu et al., 2023; Bhagat et al., 2017; Brundage & Singh, 2023; Çelikkanlı & Kızılcık, 2022; Lemmer et al., 2020; Stefanou et al., 2023; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Temiz & Yavuz, 2014; Zhao et al., 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Assimi et al., 2022; Bani-Salameh, 2017; Kaniawati et al., 2019; Kusairi et al., 2022;

Commented [A1]: Ijinkan kami menginformasikan kepada Author bahwa kebijakan terbaru OJS kami, dalam saru artikel minimal berasal dari 2 institusi yang berbeda. Pada artikel Anda hanya menyantumkan 1 institusi yang sama. Kami menyarankan untuk menambahkan author yang berasal dari institusi yang berbeda dari institusi yang sudah tertera

Commented [A2]: ?

Commented [A3]: Urutkan referensi dari tahun lampau sampai tahun terbaru

Commented [A4]: Perhatikan kembali tatacara mensitasi, sesuaikan dengan panduan manuskrip JPFI
Aykutlu *et al.,* ,menjadi Aykutlu, Ensari, & Bayrak, 2023

Cara mensitasi:
Karya dengan Dua Pengarang
Research by Wegener and Petty (1994) supports... atau (Wegener & Petty, 1994)

Karya Tiga Sampai Lima
(Kernis, Cornell, Sun, Berry, & Harlow, 1993) Atau Kernis, Cornell, Sun, Berry, & Harlow (1993) explain….
Dalam kutipan berikutnya, (Kernis et al., 1993) Atau Kernis et al. (1993) argued….

Enam Pengarang atau Lebih
Harris et al. (2001) argued... atau (Harris et al., 2001)

Perhatikan kembali tatacara penggunaan *et al*
et al digunakan untuk referensi 6 penulis atau lebih. Atau dapat digunakan untuk rujukan kedua dari referensi yang sama.
Misalkan pada Aykutlu karena terdiri dari 3 penulis maka perlu ditulis semua rujukannya, jika rujukan pada paragraf lain juga menggunakan referensiAykutlu maka pada referensi kedua dapat menggunakan *et al.,*

Rismaningsih & Nurhafsari, 2022; Suhendi & Ardiansyah, 2021; Sukarelawan, Puspitasari, Sulisworo, et al., 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Caleon & Subramaniam, 2010; Negoro & Karina, 2019; Somroob & Wattanakasiwich, 2017). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Assimi et al., 2022; Lim & Poo, 2021; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar et al., 2020; Marista et al., 2023), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial et al., 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, et al., 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials (4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of

Commented [A5]: Cek kembali sitasi, sesuaikan dengan masukan kami sebelumnya

Commented [A6]: Perhatikan kembali tahun referensi yang digunakan tidak boleh lebih dari 10 tahun (2014-2024) jika melebihi rentang tersebut maka kami sarankan untuk mengganti dengan referensi terbaru yang relefan dengan penelitian saudara

Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Guo et al., 2022; Sukarelawan, Jumadi, et al., 2021). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee et al., 2020).

The main instrument of this research is the four-level misconception diagnostic instrument on vibration and wave material (I-4WADI), which consists of 12 items. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna et al., 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz et al. (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity (Huddleston, 1956). Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Aiken, 1985). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Wahab & Tentama, 2020; Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the

Commented [A7]: Perhatikan kembali tahun referensi yang digunakan tidak boleh lebih dari 10 tahun (2014-2024) jika melebihi rentang tersebut maka kami sarankan untuk mengganti dengan referensi terbaru yang relefan dengan penelitian saudara test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams et al., 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Rusmansyah & Almubarak, 2020; Sumintono & Widhiarso, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

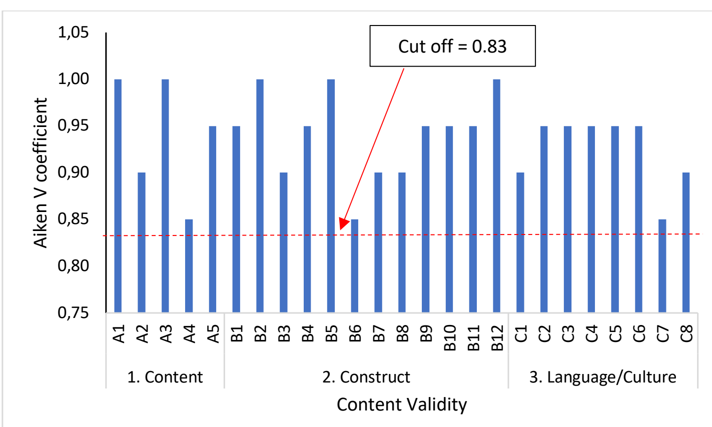

Figure 1. Aiken V validity coefficient

```
    MEASURE    Person - MAP - Item
                   <more>|<rare>
       1             .  +T
                     .  |
                     #  |
                     .  |  Q11
                     #  |  Q1
                     .  |S Q5
                     .  |  Q4
                     . T|
                        |  Q8
                     #  |
                        |M Q2     Q7
                  ####  |  Q10    Q6
       0           .## S+  Q3
                 .####  |
                  .###  |
                 .####  |S Q9
                ######  |
               .###### M|
         #############  |
             .########  |
               #######  |T
                        |
                  .### S|
                   .##  |  Q12
                        |
      -1          ####  +
                        |
                       T|
                     .  |
                        |
                        |
                     .  |
                        |
                        |
                        |
                        |
                        |
      -2             .  +
                   <less>|<freq>
     EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

d. Unidimensionality

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Hagquist et al., 2009; Kreiner, 2007; Scoulas et al., 2021; Simpelaere et al., 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, *18*(4), 515–534. https://doi.org/10.1108/ITSE-03-2020-0032

Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, *44*(1), 015701. https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, *52*(1), 14003.

Bhagat, K. K., Subheesh, N. P., Bhattacharya, B., & Chang, C.-Y. (2017). The Design and Development of Identification of Students' Misconceptions in Individualized Learning Environment (iSMILE) System. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(1), 19–34. https://doi.org/10.12973/eurasia.2017.00602a

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, *19*(2), 020112. https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, *40*(3), 313–337. https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, *19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485–493. https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, *70*(5), 717–731. https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., Karaduman, A., Swan, K., Clavé, P., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, *33*(4), 441–456. https://doi.org/10.1007/s00455-017-9873-4

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, *2022*(1), 1–16. https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, *11*(1), 16. https://doi.org/10.1186/s40468-021-00132-7

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International Journal of Nursing Studies*, *46*(3), 380–393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Commented [A8]: Pada referensi yang diblock kuning adalah referensi yang lebih 10 tahun, maka perlu diubah referensi terbaru yang relefan dengan penelitian anda

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, *9*(4), 540–553. https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, *17*(9), e0271992. https://doi.org/10.1371/journal.pone.0271992

Huddleston, E. M. (1956). Test Development on the Basis of Content Validity. *Educational and Psychological Measurement*, *16*(3), 283–293. https://doi.org/10.1177/001316445601600302

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

Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology*, *59*(3), 268–298. https://doi.org/10.1027/1901-2276.59.3.268

Kusairi, S., Sutopo, .., & Suryadi, A. (2022). Utilizing Isomorphic Multiple-Choice Items to Diagnose Students' Misconceptions in Force and Motion. *New Physics: Sae Mulli*, *72*(4), 311–318. https://doi.org/10.3938/NPSM.72.311

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, *50*(2), 375–392. https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., Zhang, W., & Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, *in press*. https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, *10*(2), 241–252. https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, *11*(August), 1–18. https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, *14*(2). https://doi.org/10.30880/jtet.2022.14.02.005

Marista, J., Latifah, S., Saregar, A., Anwar, C., Jamaluddin, W., & Fitri, M. R. (2023). *Physics E-poster to reduce the risk of misconceptions on temperature and heat: Students' perception*. 020001. https://doi.org/10.1063/5.0124209

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, *41*(2), 229–236. https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., Ceroni, L., Sonnery-Cottet, B., & Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, *23*(1), 1–8. https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, *5*(2), 69–76. https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, *in press*. https://doi.org/10.47626/2237-6089-2023-0630

Rismaningsih, F., & Nurhafsari, A. (2022). Identify Hydrostatic Misconceptions Using Four Tier Diagnostic Tests With the Help of iSpring Suite 9. *Jurnal Penelitian Pendidikan IPA*, *8*(2), 773–780. https://doi.org/10.29303/jppipa.v8i2.1290

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, *5*(3), 222. https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, *43*(1), 101071. https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, *15*(1), 66. https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, *8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, *901*, 012123. https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, *31*(1), 127–157. https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, *1918*(2). https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, *10*(4), 486–495. https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., Indratno, T. K., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, *10*(2), 216–224. https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, *55*(1), 127–141. https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, *5*(1), 1–8. https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, *88*(2), 211–221. https://doi.org/10.2298/JSC220117073S

Temiz, B. K., & Yavuz, A. (2014). Students' Misconceptions about Newton's Second Law in Outer Space. *European Journal of Physics*, *35*(4), 1–16.

Wahab, F. I. A., & Tentama, F. (2020). Construct validity and reliability test on burnout. *International Journal of Scientific and Technology Research*, *9*(1), 359–363.

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, *10*(1), 35–47. https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, *57*(1), 51–67. https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, *17*(1), 1–12. https://doi.org/10.1186/s12955-019-1168-z

7. Second revised version submitted

26 January 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K Indratno[2]*, Nurul Syafiqah Yap Abdullah[3]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Jabatan Fizik
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop misconception-diagnostic instruments. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to vibration and wave materials. The adaptation process was conducted with 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. The expert validation results were analyzed using the Aiken V technique. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about vibrations and waves.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Vibrations and waves

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a

deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials

(4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-level misconception diagnostic instrument on vibration and wave material (I-4WADI), which consists of 12 items. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, and Hambleton (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

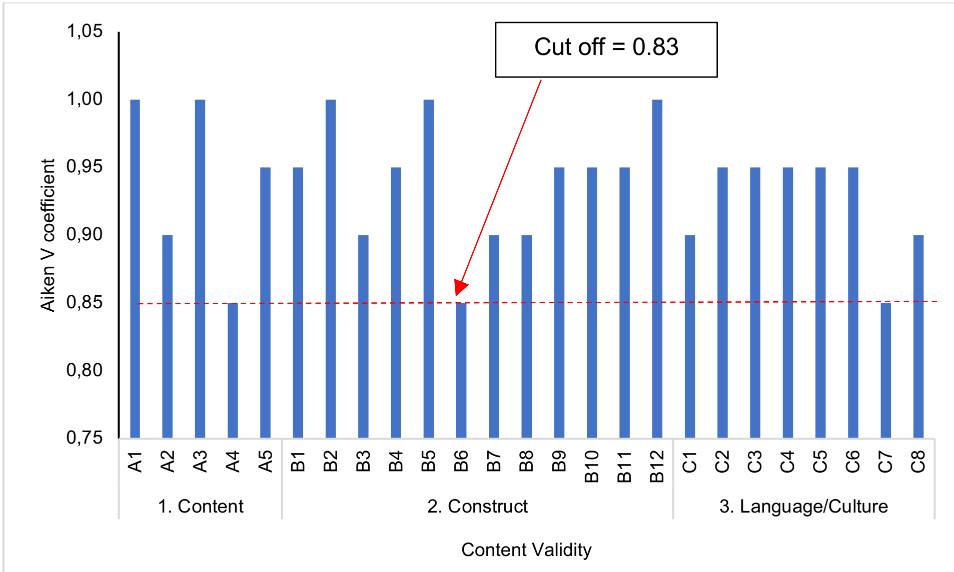

Figure 1. Aiken V validity coefficient

```
    MEASURE     Person - MAP - Item
                       <more>|<rare>
        1                  .  +T
                           .  |
                           #  |
                           .  |   Q11
                           #  |   Q1
                           .  |S  Q5
                           .  |   Q4
                           .  T|
                              |   Q8
                           #  |
                              |M  Q2      Q7
                        ####  |   Q10     Q6
        0                .## S+   Q3
                      .####   |
                       .###   |
                      .####   |S  Q9
                      ######  |
                     .###### M|
               #############  |
                   .########  |
                     #######  |T
                              |
                        .### S|
                         .##  |   Q12
                              |
       -1               ####  +
                              |
                             T|
                           .  |
                              |
                              |
                           .  |
                              |
                              |
                              |
```

```
                            |
                            |
    -2                 .   +
                      <less>|<freq>
   EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

d. Unidimensionality

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument

measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

8. Third revisions, review from Reviewer A

05 February 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K Indratno[2]*, Nurul Syafiqah Yap Abdullah[3]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Jabatan Fizik
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop misconception-diagnostic instruments. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to vibration and wave materials. The adaptation process was conducted with 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. The expert validation results were analyzed using the Aiken V technique. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about vibrations and waves.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Vibrations and waves

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a

deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials

(4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-level misconception diagnostic instrument on vibration and wave material (I-4WADI), which consists of 12 items. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

Commented [A1]: &

Commented [A2]: The method section is written with the length of 15-20% of the article length containing research design, data collection technique, data source and data analysis method.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, and Hambleton (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

Commented [A3]: Close discussions with tables or figures so readers don't look around

Commented [A4]: &... see selingkung style

Commented [A5]: You don't need to read the table anymore

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

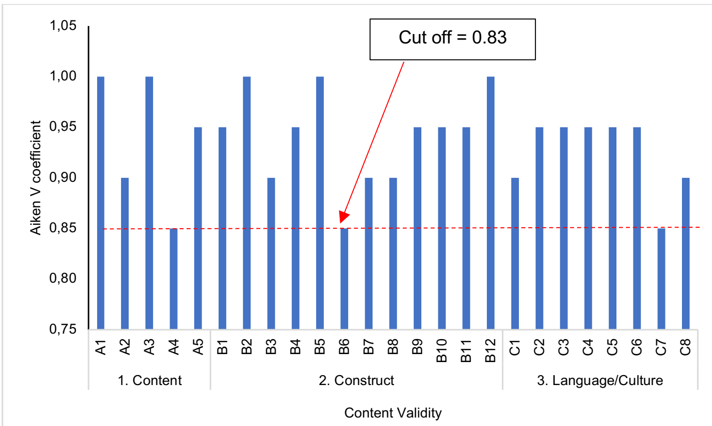

Figure 1. Aiken V validity coefficient

```
    MEASURE     Person - MAP - Item
                    <more>|<rare>
       1               .  +T
                       .  |
                       #  |
                       .  |  Q11
                       #  |  Q1
                       .  |S Q5
                       .  |  Q4
                       . T|
                          |  Q8
                       #  |
                          |M Q2      Q7
                    ####  |  Q10     Q6
       0             .## S+  Q3
                  .####   |
                   .###   |
                  .####   |S Q9
                 ######   |
                .###### M|
          #############   |
               .########  |
                #######   |T
                          |
                   .### S|
                    .##   |  Q12
                          |
      -1            ####  +
                          |
                         T|
                       .  |
                          |
                          |
                       .  |
                          |
                          |
                          |
```

```
                          |
                          |
      -2              .   +
                   <less>|<freq>
    EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

d. Unidimensionality

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument

measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

9. Third revisions, review from Reviewer B

15 February 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K Indratno[2]*, N.S.Y. Abdullah[3]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Jabatan Fizik
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop misconception-diagnostic instruments. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to vibration and wave materials. The adaptation process was conducted with 306 state high school students taking vibration and wave classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. The expert validation results were analyzed using the Aiken V technique. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about vibrations and waves.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Vibrations and waves

Commented [A1]: Please explain about the four tier diagnostic test in this part.

Commented [A2]: Better if consistency with the title

Commented [A3]: Consistency
Please explain this

Commented [A4]: Please mention how the result according to readability test.

Commented [A5]: Using Aiken V for evaluate or determine the validily of the instrument, please mention how many rater and item was used.

Commented [A6]: Mention the value of the validity and reliability

INTRODUCTION

Commented [A7]: Need emphasize the research gap for this study to make clear the situation analysis of this research.

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a

deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of vibrations and waves (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of vibration and waves have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding vibrations and waves is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of vibration and waves. Understanding electromagnetic vibrations and waves helps students understand how radio, television, internet, and mobile phone signals work. Vibrations and waves also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic vibrations and waves are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of vibrations and waves is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for vibration and wave materials

(4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to adapt a misconception diagnostic instrument on vibration and wave material in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-level misconception diagnostic instrument on vibration and wave material (I-4WADI), which consists of 12 items. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure carried out was adapting the instrument to the Indonesian cultural context. The researcher carried out the language transfer and then consulted with linguists to ensure that there was equality between the original version and the Indonesian version (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared worthy of being an expert, I-4WADI conducted a readability test on 35 students who had attended vibration and wave classes. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students completing the diagnostic process take between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

Commented [A10]: For the readability test using gap test or what?

Commented [A8]: For the aim need some revision to make it clear what the study done.

Commented [A9]: Please mention at the abstract too

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-level diagnostic instrument on vibration and wave material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, and Hambleton (2016). A summary of content validity is shown in Figure 1.

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in vibration and wave materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The results of the readability test are shown in Table 1.

Tabel 1. MRS application Readability test results

Assessment aspect	Average value	Category
Work instructions	3.37	Very good
Content Aspect	3.45	Very good
Linguistic Aspect	3.43	Very good
Graphic Aspects	3.37	Very good

The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI. Table 1 is the result of evaluating the readability of the four-level diagnostic instruments used in vibration and wave material. This evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Second, in terms of content, the average value reaches 3.45. This shows that the material presented in the diagnostic instrument is very relevant and comprehensive. This assessment also falls into the "Very Good" category, indicating that the contents of this instrument really support students' understanding of vibration and wave material. Third, in the linguistic aspect, this instrument gets an average rating of 3.43, which is also included in the "Very Good" category. This indicates that the language used in the instrument is very precise and easy to understand by students. Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instrument are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on vibration and wave materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

Commented [A11]: Need more explanation how this instrument can diagnostic the misconception on the wave and vibration material.

And I suggest for the topic is vibration and wave

Commented [A12]: Better if you add the explanation why the readability result is very good for the each the assessment aspect was measured.

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 2. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019). The item reliability value in Table 2 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the MnSq infit and outfit values were determined. Table 3 details the infit and outfit MnSq values of each item. The MnSq infit and outfit values provide information about the item's fit to the Rasch model. In this study, the MnSq infit values ranged from 0.81 to 1.40, and the MnSq outfit values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The Zstd outfit value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

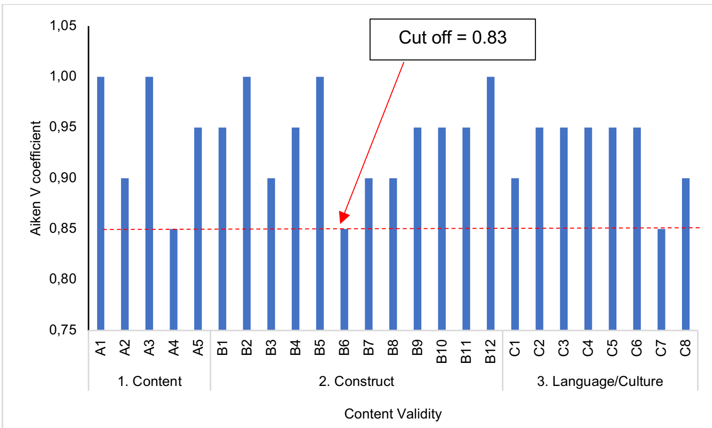

Figure 1. Aiken V validity coefficient

```
MEASURE    Person - MAP - Item
               <more>|<rare>
    1              .  +T
                   .  |
                   #  |
                   .  |  Q11
                   #  |  Q1
                   .  |S Q5
                   .  |  Q4
                   . T|
                      |  Q8
                   #  |
                      |M Q2      Q7
                ####  |  Q10     Q6
    0            .## S+  Q3
               .####  |
                .###  |
               .####  |S Q9
              ######  |
             .###### M|
       #############  |
           .########  |
             #######  |T
                      |
                .### S|
                 .##  |  Q12
                      |
   -1           ####  +
                      |
                     T|
                   .  |
                      |
                      |
                   .  |
                      |
                      |
```

```
                        |
                        |
   -2                 . +
                  <less>|<freq>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Commented [A13]: Please make sure the image is not cropped

Table 2. Summary of I-4WADI statistics

SUMMARY OF 12 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD
MEAN	881.3	306.0	.00	.06	1.01	-.08	.99	-.26
SEM	41.3	.0	.12	.00	.06	.69	.06	.58
P.SD	137.1	.0	.41	.01	.19	2.29	.19	1.93
S.SD	143.2	.0	.43	.01	.20	2.40	.20	2.02
MAX.	1252.0	306.0	.56	.08	1.40	4.45	1.36	3.65
MIN.	730.0	306.0	-1.03	.05	.76	-3.53	.71	-3.03

REAL RMSE .06 TRUE SD .41 SEPARATION 6.46 Item RELIABILITY .98
MODEL RMSE .06 TRUE SD .41 SEPARATION 6.76 Item RELIABILITY .98
S.E. OF Item MEAN = .12

Table 3. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 4. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

d. Unidimensionality

Table 4 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument

measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

CONCLUSION

Commented [A14]: Conclusion just directly state how the purpose of this research is success or not. Support with data and short analysis.

A four-level diagnostic instrument on vibration and wave materials has been developed in the context of Singaporean culture. Meanwhile, there is an urgent need for use in the Indonesian cultural context. The instrument development process requires special abilities and skills, so the cross-cultural adaptation process is one of the solutions. The results of the adaptation of 4WADI to the Indonesian cultural context have been carried out, and the results show that the four-level type diagnostic instrument on vibration and wave material is in accordance with the cultural context of high school students in Indonesia.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

10. Third revised version submitted

08 March 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K Indratno[2]*, N.S.Y. Abdullah[3], R. Oktova[4]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Jabatan Fizik
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

[4]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
r.oktova@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop diagnostic instrument for misconception. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to wave and vibration materials. This study used a cross-sectional quantitative survey method. The adaptation process was conducted with 306 state high school students taking wave and vibration classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI (Four-tier Wave and vibration Diagnostic Instrument) was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. As a result, the I-4WADI has excellent readability. The validation results of six experts were analyzed using the Aiken V technique. A total of 25 items were used for expert validation. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. The infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about wave and vibration.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Wave and vibration

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of wave and vibration (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of wave and vibration have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding wave and vibration is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of wave and vibration. Understanding electromagnetic wave and vibration helps students understand how radio, television, internet, and mobile phone signals work. Wave and vibration also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic wave and vibration are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of wave and vibration is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for wave and vibration materials (4WADI) was previously carried out by Caleon & Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to develop a diagnostic instrument for misconceptions regarding waves and vibrations through an adaptation process in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-tier wave and vibration diagnostic instrument for misconception (4WADI), which consists of 12 items. As the name suggests, each question in the diagnostic test instrument has four levels of questions. The first level is conceptual questions. The second level is the respondent's level of confidence in answering the first level. The third level is a choice of explanation of the phenomenon at the first level. The fourth level is the level of confidence in the choices given at the third level. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure was to adapt the instrument to the Indonesian cultural context. The researcher translated the language and then consulted with a linguist to ensure equality between the original and Indonesian versions (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared suitable by experts, I-4WADI was tested for readability on 35 students who had taken the wave and vibration class. Readability tests were analyzed using scoring techniques. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students complete the diagnostic process in between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-tier diagnostic instrument on wave and vibration material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, & Hambleton (2016). A summary of content validity is shown in Figure 1.

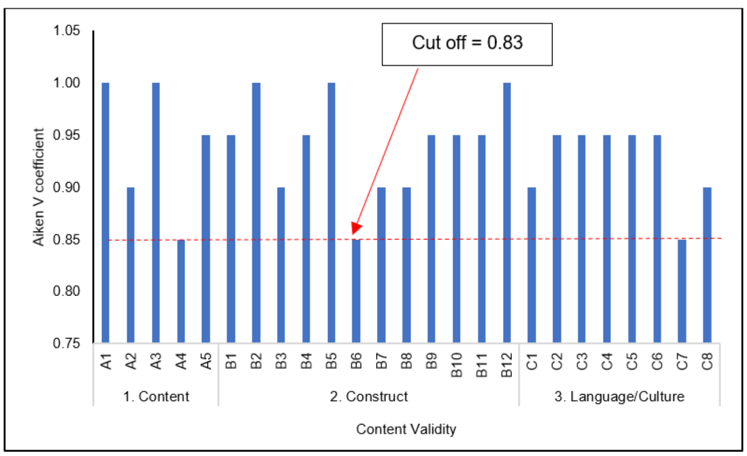

Figure 1. Aiken V validity coefficient

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in wave and vibration materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI.

Readability evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Before entering the diagnostic section, the initial part of the I-4WADI consists of instructions for use and respondent demographics. The information contained in the I-4WADI consists of filling requirements and a brief explanation of the question model used in the instrument. The instructions section of the misconception diagnostic instrument is very important for potential respondents because it provides clear directions on how to answer the questions correctly. Good instructions can help potential respondents understand the purpose of the instrument and how they should interpret the questions and provide accurate responses. This can also help reduce misunderstandings or wrong interpretations that can interfere with the validity of measurement results.

Second, in terms of content, the average value reached 3.45. This shows that the material presented on the diagnostic instrument is very relevant and comprehensive. This assessment is also included in the "Very Good" category, which shows that the content of this instrument really supports students' understanding of waves and vibrations. This cannot be separated from the quality of 4WADI, as Caleon & Subramaniam (2010) reported. They have undergone a series of structured and systematic steps in producing 4WADI, starting from Content Validation and Piloting, Construction, Administration, and Validation.

Third, in the linguistic aspect, I-4WADI obtained an average score of 3.43, which is also included in the "Very Good" category. This shows that the language used is very precise and easy for students to understand. Linguists have guaranteed equivalence between the original language version of the instrument (4-WADI) and the Indonesian version (I-4WADI) through a forward-backward translation procedure. They are responsible for ensuring that the translation reflects the original text's exact meaning and takes into account the cultural and linguistic context of the target language. Linguist experts provide recommendations and ensure that the target Indonesian students can understand the translated instruments well. More than that, linguists also recommend some relevant vocabulary, avoiding phrases or terms that are ambiguous or difficult for potential respondents to understand.

Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instruments are very good and support students' experience in answering questions..

Thus, the readability results of the four-level diagnostic instrument on wave and vibration materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 1.

Table 1. Summary of I-4WADI statistics

Index of item	Score
Infit MnSq	1.01
Infit ZStd	-0.08
Outfit MnSq	0.99
Outfit ZStd	-0.26
Separation	6.46
Reliability	0.98

Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019). The item reliability value in Table 1 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. **Item Difficulty Distribution (LVI and Wright map)**

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. **Item fit**

To obtain information on item suitability for the Rasch modeling, the infit and outfit MnSq values were determined. Table 2 details the infit and outfit MnSq values of each item. The infit and outfit MnSq values provide information about the item's fit to the Rasch model. In this study, the infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The outfit Zstd value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

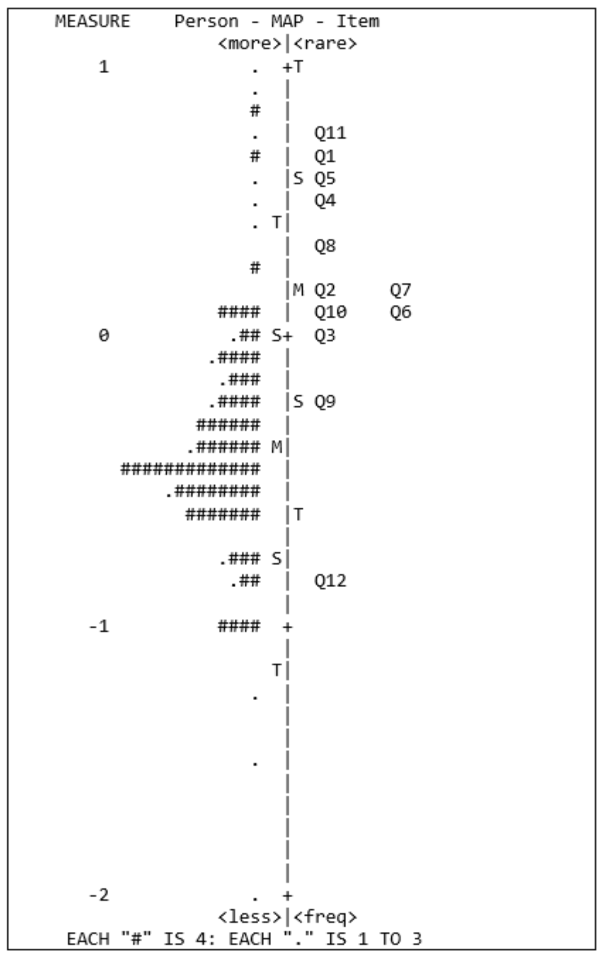

Figure 2. Distribution of items in the person-item map (Wright map)

d. Unidimensionality

Table 3 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

Table 2. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 3. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

CONCLUSION

A four-level diagnostic instrument on wave and vibration materials has been developed through an adaptation process in the Indonesian cultural context. The results of adapting 4WADI to the Indonesian cultural context have been carried out into I-4WADI, and the results show that the four-level type diagnostic instrument on wave and vibration material is appropriate to the cultural context of high school students in Indonesia—the results of the expert assessment of content validity range from 0.85 to 1.00. The readability level is very good, with an average score of 3.37. Empirically, I-4WADI has infit MnSq values ranging from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98.

Therefore, I-4WADI meets the criteria for validity and reliability.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-

021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

11. Forth revisions, review from Reviewer B

25 March 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K Indratno[2]*, N.S.Y. Abdullah[3], R. Oktova[4]

[1]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Jabatan Fizik
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

[4]Department of Physics Education
Universitas Ahmad Dahlan, Indonesia
r.oktova@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop diagnostic instrument for misconception. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to wave and vibration materials. This study used a cross-sectional quantitative survey method. The adaptation process was conducted with 306 state high school students taking wave and vibration classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI (Four-tier Wave and vibration Diagnostic Instrument) was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI was tested for readability on 35 students. As a result, the I-4WADI has excellent readability. The validation results of six experts were analyzed using the Aiken V technique. A total of 25 items were used for expert validation. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. The infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about wave and vibration.

Commented [A1]: Explain this.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Wave and vibration

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of wave and vibration (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of wave and vibration have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding wave and vibration is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of wave and vibration. Understanding electromagnetic wave and vibration helps students understand how radio, television, internet, and mobile phone signals work. Wave and vibration also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic wave and vibration are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of wave and vibration is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is

Commented [A2]: Need emphasize the research gap for this study to make clear the situation analysis of this research. For eth example is how the situation of the diagnostic test and also why need use the Rasch model to evaluate?

an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for wave and vibration materials (4WADI) was previously carried out by Caleon & Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). So, another approach is needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Therefore, this research aims to develop a diagnostic instrument for misconceptions regarding waves and vibrations through an adaptation process in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-tier wave and vibration diagnostic instrument for misconception (4WADI), which consists of 12 items. As the name suggests, each question in the diagnostic test instrument has four levels of questions. The first level is conceptual questions. The second level is the respondent's level of confidence in answering the first level. The third level is a choice of explanation of the phenomenon at the first level. The fourth level is the level of confidence in the choices given at the third level. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure was to adapt the instrument to the Indonesian cultural context. The researcher translated the language and then consulted with a linguist to ensure equality between the original and Indonesian versions (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared suitable by experts, I-4WADI was tested for readability on 35 students who had taken the wave and vibration class. Readability tests were analyzed using scoring techniques. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students complete the diagnostic process in between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULTS AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-tier diagnostic instrument on wave and vibration material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, & Hambleton (2016). A summary of content validity is shown in Figure 1.

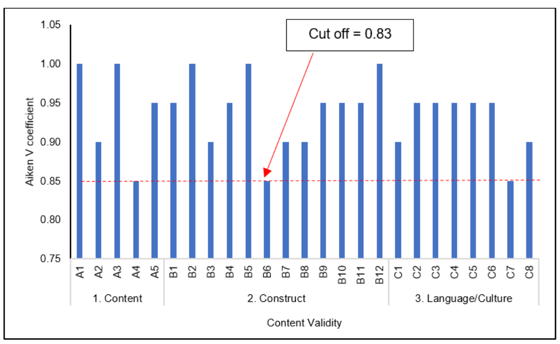

Figure 1. Aiken V validity coefficient

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in wave and vibration materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI.

Readability evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Before entering the diagnostic section, the initial part of the I-4WADI consists of instructions for use and respondent demographics. The information contained in the I-4WADI consists of filling requirements and a brief explanation of the question model used in the instrument. The instructions section of the misconception diagnostic instrument is very important for potential respondents because it provides clear directions on how to answer the questions correctly. Good instructions can help potential respondents understand the purpose of the instrument and how they should interpret the questions and provide accurate responses. This can also help reduce misunderstandings or wrong interpretations that can interfere with the validity of measurement results.

Second, in terms of content, the average value reached 3.45. This shows that the material presented on the diagnostic instrument is very relevant and comprehensive. This assessment is also included in the "Very Good" category, which shows that the content of this instrument really supports students' understanding of waves and vibrations. This cannot be separated from the quality of 4WADI, as Caleon & Subramaniam (2010) reported. They have undergone a series of structured and systematic steps in producing 4WADI, starting from Content Validation and Piloting, Construction, Administration, and Validation.

Third, in the linguistic aspect, I-4WADI obtained an average score of 3.43, which is also included in the "Very Good" category. This shows that the language used is very precise and easy for students to understand. Linguists have guaranteed equivalence between the original language version of the instrument (4-WADI) and the Indonesian version (I-4WADI) through a forward-backward translation procedure. They are responsible for ensuring that the translation reflects the original text's exact meaning and takes into account the cultural and linguistic context of the target language. Linguist experts provide recommendations and ensure that the target Indonesian students can understand the translated instruments well. More than that, linguists also recommend some relevant vocabulary, avoiding phrases or terms that are ambiguous or difficult for potential respondents to understand.

Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instruments are very good and support students' experience in answering questions..

Thus, the readability results of the four-level diagnostic instrument on wave and vibration materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 1.

Table 1. Summary of I-4WADI statistics

Index of item	Score
Infit MnSq	1.01
Infit ZStd	-0.08
Outfit MnSq	0.99
Outfit ZStd	-0.26
Separation	6.46
Reliability	0.98

Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019). The item reliability value in Table 1 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the infit and outfit MnSq values were determined. Table 2 details the infit and outfit MnSq values of each item. The infit and outfit MnSq values provide information about the item's fit to the Rasch model. In this study, the infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The outfit Zstd value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

```
        MEASURE    Person - MAP - Item
                         <more>|<rare>
             1              .  +T
                            .  |
                            #  |
                            .  |  Q11
                            #  |  Q1
                            .  |S Q5
                            .  |  Q4
                            . T|
                               |  Q8
                            #  |
                               |M Q2      Q7
                         ####  |  Q10     Q6
             0            .## S+  Q3
                        .####  |
                         .###  |
                        .####  |S Q9
                       ######  |
                      .###### M|
                ############# |
                    .########  |
                      #######  |T
                               |
                         .### S|
                          .##  |  Q12
                               |
            -1           ####  +
                               |
                              T|
                            .  |
                               |
                               |
                            .  |
                               |
                               |
                               |
                               |
                               |
            -2              .  +
                         <less>|<freq>
          EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

d. Unidimensionality

Table 3 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

Table 2. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 3. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

CONCLUSION

A four-level diagnostic instrument on wave and vibration materials has been developed through an adaptation process in the Indonesian cultural context. The results of adapting 4WADI to the Indonesian cultural context have been carried out into I-4WADI, and the results show that the four-level type diagnostic instrument on wave and vibration material is appropriate to the cultural context of high school students in Indonesia—the results of the expert assessment of content validity range from 0.85 to 1.00. The readability level is very good, with an average score of 3.37. Empirically, I-4WADI has infit MnSq values ranging from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98.

Therefore, I-4WADI meets the criteria for validity and reliability.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-

021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

12. Forth revised version submitted

27 March 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
mm 20XX

Jurnal Pendidikan Fisika Indonesia xx (x) (20xx) x-x
DOI: xx.xxxxx/jpfi.xxxxx.xxxx

http://journal.unnes.ac.id/nju/index.php/jpfi

QUALITY OF FOUR-TIER DIAGNOSTIC TEST ON WAVE AND VIBRATION MATERIALS: AN EMPIRICAL STUDY USING RASCH MODELING

M. I. Sukarelawan[1], T. K. Indratno[2]*, N.S.Y. Abdullah[3], R. Oktova[4]

[1]Physics Education Study Program
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id

[2] Physics Education Study Program
Universitas Ahmad Dahlan, Indonesia
tonikus@staff.uad.ac.id

[3]Department of Physics
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

[4]Physics Education Study Program
Universitas Ahmad Dahlan, Indonesia
r.oktova@pfis.uad.ac.id

Received: dd mm yyyy. Accepted: dd mm yyyy. Published: mm yyyy

Abstract

Not all teachers have the ability to develop diagnostic instrument for misconception. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to wave and vibration materials. This study used a cross-sectional quantitative survey method. The adaptation process was conducted with 306 state high school students taking wave and vibration classes. Respondents were selected using convenience sampling techniques. To produce an instrument appropriate to the Indonesian cultural context, 4WADI (Four-tier Wave and vibration Diagnostic Instrument) was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts. Before the empirical test was carried out, I-4WADI (Indonesian Four-tier Wave and vibration Diagnostic Instrument) was tested for readability on 35 students. As a result, the I-4WADI has excellent readability. The validation results of six experts were analyzed using the Aiken V technique. A total of 25 items were used for expert validation. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. The infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about wave and vibration.

Key words: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Wave and vibration

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of wave and vibration (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of wave and vibration have significantly impacted a basic understanding of the physical world and have relevance in many aspects of everyday life and other sciences. Understanding wave and vibration is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of wave and vibration. Understanding electromagnetic wave and vibration helps students understand how radio, television, internet, and mobile phone signals work. Wave and vibration also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. The concepts of electromagnetic wave and vibration are also relevant to the understanding of the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, an understanding of wave and vibration is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi et al., 2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, et al., 2022; Zhao et al., 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although there has yet to be an agreement from experts, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for wave and vibration materials (4WADI) was previously carried out by Caleon & Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). The difficulty level of items in CTT depends on the sample or the respondent's ability (Caycho-Rodríguez et al., 2022). This results in a person's score being low if the question is difficult and high if the question is easy. Therefore, other approaches are needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Rasch modeling can be an integrated approach to provide more comprehensive information regarding the psychometric properties of 4WADI that have not yet been reported (Sukerelawan, Jumadi, Kuswanto, & Thohir, 2021). Through Rasch modeling, the psychometric properties of 4WADI in Indonesian culture can be evaluated at the individual item level. The use of Rasch modeling to evaluate the psychometric properties of 4WADI in various cultures has not been widely reported, including in Indonesia. Therefore, this research aims to develop a diagnostic instrument for misconceptions regarding waves and vibrations through an adaptation process in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from class XI and XII from 3 public schools. The sample size used has taken into account the accuracy of measurement to 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Ling Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-tier wave and vibration diagnostic instrument for misconception (4WADI), which consists of 12 items. As the name suggests, each question in the diagnostic test instrument has four levels of questions. The first level is conceptual questions. The second level is the respondent's level of confidence in answering the first level. The third level is a choice of explanation of the phenomenon at the first level. The fourth level is the level of confidence in the choices given at the third level. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure was to adapt the instrument to the Indonesian cultural context. The researcher translated the language and then consulted with a linguist to ensure equality between the original and Indonesian versions (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared suitable by experts, I-4WADI was tested for readability on 35 students who had taken the wave and vibration class. Readability tests were analyzed using scoring techniques. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students complete the diagnostic process in between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan, Jumadi, et al., 2021; Sukarelawan, Puspitasari, Rahmatika, et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variance explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULT AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-tier diagnostic instrument on wave and vibration material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, & Hambleton (2016). A summary of content validity is shown in Figure 1.

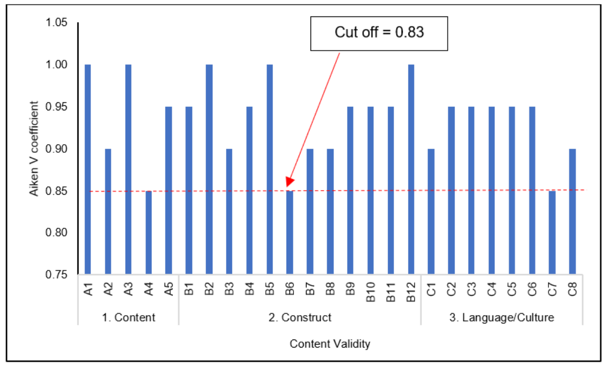

Figure 1. Aiken V validity coefficient

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument succeeded in meeting the established standards, 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in wave and vibration materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI.

Readability evaluation refers to four important aspects, namely, work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Before entering the diagnostic section, the initial part of the I-4WADI consists of instructions for use and respondent demographics. The information contained in the I-4WADI consists of filling requirements and a brief explanation of the question model used in the instrument. The instructions section of the misconception diagnostic instrument is very important for potential respondents because it provides clear directions on how to answer the questions correctly. Good instructions can help potential respondents understand the purpose of the instrument and how they should interpret the questions and provide accurate responses. This can also help reduce misunderstandings or wrong interpretations that can interfere with the validity of measurement results.

Second, in terms of content, the average value reached 3.45. This shows that the material presented on the diagnostic instrument is very relevant and comprehensive. This assessment is also included in the "Very Good" category, which shows that the content of this instrument really supports students' understanding of waves and vibrations. This cannot be separated from the quality of 4WADI, as Caleon & Subramaniam (2010) reported. They have undergone a series of structured and systematic steps in producing 4WADI, starting from Content Validation and Piloting, Construction, Administration, and Validation.

Third, in the linguistic aspect, I-4WADI obtained an average score of 3.43, which is also included in the "Very Good" category. This shows that the language used is very precise and easy for students to understand. Linguists have guaranteed equivalence between the original language version of the instrument (4-WADI) and the Indonesian version (I-4WADI) through a forward-backward translation procedure. They are responsible for ensuring that the translation reflects the original text's exact meaning and takes into account the cultural and linguistic context of the target language. Linguist experts provide recommendations and ensure that the target Indonesian students can understand the translated instruments well. More than that, linguists also recommend some relevant vocabulary, avoiding phrases or terms that are ambiguous or difficult for potential respondents to understand.

Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instruments are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on wave and vibration materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 1. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019).

Table 1. Summary of I-4WADI statistics

Index of item	Score
Infit MnSq	1.01
Infit ZStd	-0.08
Outfit MnSq	0.99
Outfit ZStd	-0.26
Separation	6.46
Reliability	0.98

The item reliability value in Table 1 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the infit and outfit MnSq values were determined. Table 2 details the infit and outfit MnSq values of each item. The infit and outfit MnSq values provide information about the item's fit to the Rasch model. In this study, the infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The outfit Zstd value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that there may be some items that do not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

d. Unidimensionality

Table 2 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variance explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variance explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

```
MEASURE    Person - MAP - Item
               <more>|<rare>
   1               .  +T
                   .  |
                   #  |
                   .  |  Q11
                   #  |  Q1
                   .  |S Q5
                   .  |  Q4
                   . T|
                      |  Q8
                   #  |
                      |M Q2     Q7
                ####  |  Q10    Q6
   0             .## S+  Q3
               .####  |
                .###  |
               .####  |S Q9
              ######  |
             .###### M|
       #############  |
           .########  |
             #######  |T
                      |
                .### S|
                 .##  |  Q12
                      |
  -1            ####  +
                      |
                     T|
                   .  |
                      |
                      |
                   .  |
                      |
                      |
                      |
                      |
                      |
  -2               .  +
               <less>|<freq>
 EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. Fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 3. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units

	Eigenvalue	Observed		Expected
Total raw variance in observations =	16.7485	100.0%		100.0%
Raw variance explained by measures =	4.7485	28.4%		28.4%
Raw variance explained by persons =	.5610	3.3%		3.4%
Raw Variance explained by items =	4.1875	25.0%		25.1%
Raw unexplained variance (total) =	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast =	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast =	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast =	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast =	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast =	1.1818	7.1%	9.8%	

CONCLUSION

A four-level diagnostic instrument on wave and vibration materials has been developed through an adaptation process in the Indonesian cultural context. The results of adapting 4WADI to the Indonesian cultural context have been carried out into I-4WADI, and the results show that the four-level type diagnostic instrument on wave and vibration material is appropriate to the cultural context of high school students in Indonesia—the results of the expert assessment of content validity range from 0.85 to 1.00. The readability level is very good, with an average score of 3.37. Empirically, I-4WADI has infit MnSq values ranging from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. Therefore, I-4WADI meets the criteria for validity and reliability.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati Idrissi, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Caycho-Rodríguez, T., Vilca, L. W., Cervigni, M., Gallegos, M., Martino, P., Portillo, N., … Burgos Videla, C. (2022). Fear of COVID-19 scale: Validity, reliability and factorial invariance in Argentina's general population. *Death Studies*, *46*(3), 543–552. doi:10.1080/07481187.2020.1836071

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., … Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M., … Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., & Thohir, M. A. (2021). The Indonesian Version of the Physics Metacognition Inventory: Confirmatory Factor Analysis and Rasch Model. *European Journal of Educational Research*, *10*(4), 2133–

2144. doi:10.12973/eu-jer.10.4.2133

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., … Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z

13. Paper Accepted

23 April 2024

Jurnal Pendidikan Fisika Indonesia

P-ISSN-1693-1246, E-ISSN-2355-3812 (Nationally Accredited by RISTEKDIKTI with SINTA 2 No.148/M/KPT/2020)
Physics Department, Faculty of Mathematics and Natural Sciences, Semarang State University (Unnes)
D7 Building, 2nd Floor, Sekaran Campus, Gunungpati, Semarang, Central Java, Indonesia, 50229

LETTER OF ACCEPTANCE

Number : 046/JPFI/B.IV/2024

To : **Toni Kus Indratno**
Physics Education Study Program Universitas Ahmad Dahlan,
Indonesia *e-mail : tonikus@staff.uad.ac.id

We are pleased to announce you, that the following article:

ID	:	47493
Title	:	Quality of Four-Tier Diagnostic Test on Wave and Vibration Materials: An Empirical Study Using Rasch Modeling
Author	:	**Moh. Irma Sukarelawan, Toni Kus Indratno, Nurul Syafiqah Yap Abdullah, Raden Oktova**
Decision	:	ACCEPTED

has been accepted for publication in Indonesian Journal of Physics Education Volume 20 Number 1 in 2024, based on the selection and reviewing process by the referees. The paper has also been revised by the language and layout Editor.

Thank you for your contribution to the Indonesian Journal of Physics Education and we look forward to receiving further submission from you.

Semarang, 23th April 2024

Editor in Chief

JURNAL PENDIDIKAN FISIKA INDONESIA
FISIKA FMIPA UNNES

Dr.PratiwiDwijananti,M.Si

14. Paper published

01 July 2024

P-ISSN: 1693-1246
E-ISSN: 2355-3812
June 2024

Jurnal Pendidikan Fisika Indonesia 20 (1) (2024) 27-37
DOI: 10.15294/jpfi.v20i1.47493

http://journal.unnes.ac.id/nju/index.php/jpfi

Quality of Four-Tier Diagnostic Test on Wave and Vibration Materials: An Empirical Study Using Rasch Modeling

Moh. Irma Sukarelawan[1], Toni Kus Indratno[1*], Raden Oktova[1], Nurul Syafiqah Yap Abdullah[2]

[1]Physics Education Study Program,
Universitas Ahmad Dahlan, Indonesia
irma.sukarelawan@pfis.uad.ac.id, tonikus@staff.uad.ac.id, r.oktova@pfis.uad.ac.id

[3]Department of Physics,
Universiti Pendidikan Sultan Idris, Malaysia
syafiqah @fsmt.upsi.edu.my

Received: 21 February 2024. Accepted: 19 May 2024. Published: 25 June 2024

Abstract

Not all teachers have the ability to develop diagnostic instruments for misconception. Meanwhile, there is an urgent need to diagnose student misconceptions. Therefore, this research aims to carry out the process of adapting misconception-diagnostic instruments to wave and vibration materials. This study used a cross-sectional quantitative survey method. The adaptation process was conducted with 306 state high school students taking wave and vibration classes. Respondents were selected using convenience sampling techniques. 4WADI (Four-tier Wave and vibration Diagnostic Instrument) was translated into Indonesian, consulted with language experts, and assessed by material and evaluation experts to produce an instrument appropriate to the Indonesian cultural context. Before the empirical test was carried out, I-4WADI (Indonesian et al. and vibration Diagnostic Instrument) was tested for readability on 35 students. As a result, the I-4WADI has excellent readability. The validation results of six experts were analyzed using the Aiken V technique. A total of 25 items were used for expert validation. Empirical validation data were analyzed using the Rasch model. Review empirical validation from the aspects of item fit and unidimensionality. The analysis results show that I-4WADI has good quality based on validity and reliability. The infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. So, I-4WADI can be applied to high school students in Indonesia. These findings have practical implications for teachers who wish to diagnose students' misconceptions about waves and vibration.

Keywords: Diagnostic tests, Instrument adaptation, Misconceptions, Rasch modeling, Wave and vibration

INTRODUCTION

Understanding concepts in various fields of science is an important foundation for developing human knowledge. However, throughout the history of education, the misconception is one of the classic problems that stand out (Lemmer, Kriek, & Erasmus, 2020; Çelikkanlı & Kızılcık, 2022; Sukarelawan, Puspitasari, Sulisworo, Kuswanto, & Jumadi, 2022; Aykutlu, Ensari, & Bayrak, 2023; Brundage & Singh, 2023; Stefanou, Stylos, Georgopoulos, & Kotsis, 2023; Zhao, Zhang, Cui, Hu, & Dai, 2023). Misconception refers to a wrong or erroneous understanding of certain concepts and is believed to be accurate. Misconceptions can hinder effective learning and be detrimental to a deep understanding of the lesson material (Luce & Callanan, 2020).

Misconceptions have become a serious concern in the world of education in the last four decades, especially in the context of learning physics (Bani-Salameh, 2017; Kaniawati et al., 2019; Suhendi & Ardiansyah, 2021; Assimi, Janati Idrissi, Zerhane, & Boubih, 2022; Dwiastuti, Sukarelawan, & Sriyanto, 2022). When students develop a wrong understanding of key concepts in physics, this can interfere with their knowledge development, hinder their ability to solve problems, and even hinder their interest in studying physics.

One of the physics concepts that is important for students is the matter of wave and vibration (Somroob & Wattanakasiwich, 2017; Negoro & Karina, 2019). The concepts of wave and vibration have significantly impacted a basic understanding of the physical world and are relevant in many aspects of everyday life and other sciences. Understanding wave and vibration is the basis for understanding various other physics concepts, such as mechanics, acoustics, optics, and electromagnetism. Mobile phones, radios, computers, and medical devices such as medical imaging machines and sonar are technologies in everyday life based on the principle of wave and vibration. Understanding electromagnetic waves and vibrations helps students understand how radio, television, internet, and mobile phone signals work. Wave and vibration also have a strong connection to music and art. Understanding sound frequencies helps students understand the principles of music and acoustics. Electromagnetic waves and vibrations are also relevant to understanding the universe. For example, electromagnetic radiation from objects in space is how we study and understand galaxies, planets, and other astronomical objects. In the era of modern technology, understanding waves and vibrations is also relevant in developing technologies such as transmitting data via vibrations and electromagnetic waves (Wi-Fi, Bluetooth), facial recognition technology, and so on.

In addition to its negative impact on individual learning outcomes, misconceptions challenge teachers and educational curricula. Teachers often have to identify (Lim & Poo, 2021; Assimi, Janati, Zerhane, & Boubih ,2022; Soeharto & Csapó, 2022; Sukarelawan, Puspitasari, Sulisworo, 2022; Zhao, Zhang, Cui, Hu, & Dai 2023) and overcome misconceptions that students may have (Haidar, Yuliati, & Handayanto, 2020), while the curriculum must be carefully designed to prevent or reduce misconceptions as much as possible (Maknun & Marwiah, 2022; Syahrial, Ilmah, Yahmin, Munzil, & Muntholib, 2023). The accuracy of curriculum design to reduce and overcome misconceptions depends on diagnostic instruments as a tool for identification. Although various diagnostic instruments have been developed for this purpose, many have significant limitations. Some of these instruments may not be accurate or sensitive enough to detect certain misconceptions.

Developing a misconception diagnostic instrument is a complex task and requires special attention because it involves an in-depth understanding of the misconceptions and the ability to design effective instruments. The physics misconception diagnostic instrument must be sophisticated, relevant, and efficient to provide comprehensive information on students' understanding of concepts (Sukarelawan, Sriyanto, Puspitasari, Sulisworo, & Hikmah, 2021). In reality, not all teachers can develop misconceptions about diagnostic instruments. At the same time, there is an urgent need to use instruments to diagnose students' misconceptions.

Although experts have yet to agree, the four-level multiple choice instrument is an effective and accurate diagnostic instrument to be applied to students in large classes. Such instruments should be able to identify misconceptions appropriately, provide useful feedback to teachers and students, and be adaptable to the needs of different curricula and teaching methods.

One practical and effective instant effort is to use instruments that previous researchers have developed. However, an adaptation process is needed in different language and cultural contexts. The development of a four-level diagnostic instrument for wave and vibration materials (4WADI) was previously carried out by Caleon and Subramaniam (2010). They have reported the quality of 12 items developed in a classical test theory (CTT) perspective. By using respondents as many as 598 Singapore high school students, the 4WADI instrument has shown its performance in diagnosing student misconceptions. However, 4WADI cannot be applied in the cultural context of Indonesia because it requires an adaptation process. In addition, there is limited information that can be reported when using CTT (Sukarelawan, Jumadi, Kuswanto, Soeharto, & Hikmah, 2021; Guo, Lu, Johnson, & McCaffrey, 2022). The difficulty level of items in CTT depends on the sample or the respondent's ability (Caycho-Rodríguez et al., 2022). This results in a person's score being low if the question is difficult and high if the question is easy. Therefore, other approaches are needed, such as Rasch modeling based on modern test theory, which can be an alternative to CTT. Rasch modeling can be an integrated approach to provide more comprehensive information regarding the psychometric properties of 4WADI that have not yet been reported (Sukarelawan, Jumadi, Kuswanto, & Thohir, 2021). Through Rasch modeling, the psychometric properties of 4WADI in Indonesian culture can be evaluated at the individual item level. The use of Rasch modeling to evaluate the psychometric properties of 4WADI in various cultures has not been widely reported, including in Indonesia. Therefore, this research aims to develop a diagnostic instrument for misconceptions regarding waves and vibrations through an adaptation process in the Indonesian cultural context.

METHOD

This study used a cross-sectional quantitative survey method. The main sample used in this study was 306 public high school students selected using the convenience sampling technique. Respondents consisted of 67.6% (207 of 306 students) female and 32.4% (99 of 306 students) male. All respondents came from classes XI and XII from 3 public schools. The sample size used has considered the measurement accuracy to be 0.5 logit accuracy and a 99% confidence level (Chen et al., 2014; Lee, Chinna, & Sumintono, 2020).

The main instrument of this research is the four-tier wave and vibration diagnostic instrument for misconception (4WADI), which consists of 12 items. As the name suggests, each question in the diagnostic test instrument has four levels of questions. The first level is conceptual questions. The second level is the respondent's level of confidence in answering the first level. The third level is a choice of explanation of the phenomenon at the first level. The fourth level is the level of confidence in the choices given at the third level. The cognitive level of the questions used is C2 (understanding) to C4 (analyzing). The supporting instruments used to support the adaptation process in the Indonesian cultural context include advice sheets from language experts, expert assessment sheets, and readability test sheets for students. The language expert's suggestion sheet contains a column of suggestions and improvements descriptively. The expert assessment sheet comprises 25 items spread across material, construction, and language/cultural aspects. The readability test sheet for students consists of 12 items spread over aspects of work instructions, content aspects, linguistic aspects, and graphical aspects.

The initial procedure was to adapt the instrument to the Indonesian cultural context. The researcher translated the language and then consulted with a linguist to ensure equality between the original and Indonesian versions (I-4WADI). After going through the adaptation process, I-4WADI was reassessed by six material and evaluation experts. After being declared suitable by experts, I-4WADI was tested for readability on 35 students who had taken the wave and vibration class. Readability tests were analyzed using scoring techniques. After passing the readability test, I-4WADI was tested empirically on students. Empirical tests are carried out online to facilitate the data administration process. Students complete the diagnostic process in between 30 to 40 minutes.

The data obtained in this study are in the form of suggestions and corrections from language experts in the form of qualitative data. Expert validation data were analyzed using the Aiken V technique. Data from the readability test results were analyzed using a scale conversion guideline of 4. The psychometric properties of I-4WADI were empirically analyzed using Rasch modeling. The reported index is in the form of reliability in terms of separation and reliability aspects (Yuhanna, Al Muhdhar, Gofur, & Hassan, 2021). Then, use the Logit Value of the Item (LVI) combined with the Wright map to evaluate the distribution of the difficulty levels of the questions (Sukarelawan et al., 2021; Sukarelawan et al. et al., 2022). The fit analysis to the Rasch modeling was evaluated based on the infit and outfit MnSq (Matheny & Clanton, 2020). The acceptance limit for the model's fit is when the MnSq infit and outfit values are in the range of 0.5–1.5 (Sumintono & Widhiarso, 2014, 2015). Another index reported is unidimensionality. I-4WADI is unidimensional if the score of raw variances explained by measure exceeds 30%, and the Unexplained variance in 1st contrast does not exceed 15%.

RESULT AND DISCUSSION

1. Content Validity

Before being validated by experts, a four-tier diagnostic instrument on wave and vibration material (4WADI) was adapted using the model proposed by Muñiz, Elosua, Padilla, and Hambleton (2016). A summary of content validity is shown in Figure 1.

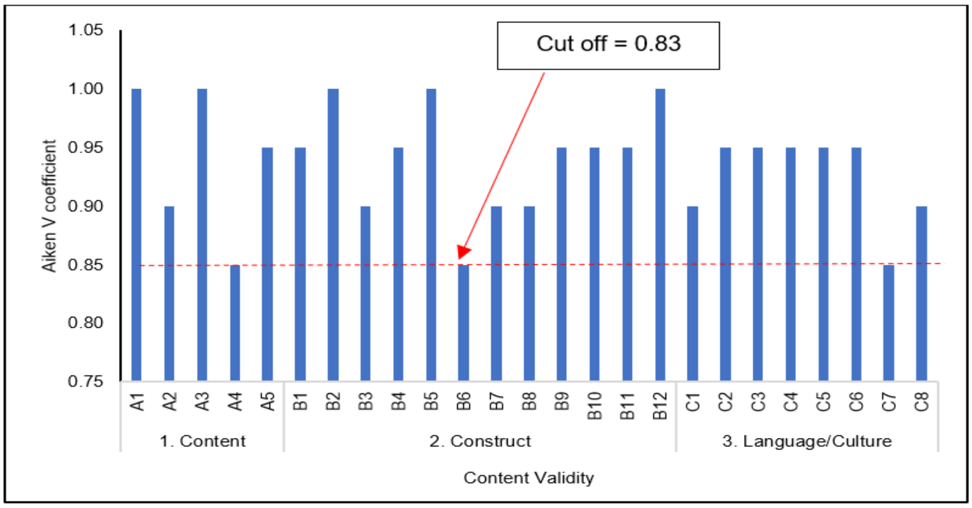

Figure 1. Aiken V validity coefficient

Content validity plays an important role in developing tests to achieve predictive validity. Validation results from a number of material experts and evaluation experts show that the Aiken V coefficient value is in a very satisfactory range, namely between 0.85 and 1.00. This indicates that the diagnostic instrument used in this research has a very high level of validity. In the context of the minimum limit value of the validity criteria used, all items in the misconception diagnostic instrument met the established standards of 0.83 (Larasati, Supahar, & Yunanta, 2020). Therefore, this instrument can be relied on to identify misconceptions in wave and vibration materials. These findings align with the results of the content validity test carried out by (Caleon & Subramaniam, 2010).

2. Respondent readability test results

After being translated into I-4WADI and declared valid by experts, I-4WADI was tested for readability on 35 students. The readability test is an important process in I-4WADI adaptation. The readability test helps measure the extent to which students can read and understand the instrument easily. By identifying potential ambiguity, confusion, or difficult language, readability tests can help improve the accuracy of I-4WADI.

Readability evaluation refers to four important aspects: work instructions, content, language, and graphics. The results of this assessment are very satisfying. First, in the aspect of work instructions, this diagnostic instrument obtained an average score of 3.37. This shows that the instructions given to students are very good and easy to understand. This assessment falls into the "Very Good" category, indicating that the instrument provides clear directions for students in carrying out their assignments.

Before entering the diagnostic section, the initial part of the I-4WADI consists of instructions for use and respondent demographics. The information contained in the I-4WADI consists of filling requirements and a brief explanation of the question model used in the instrument. The instructions section of the misconception diagnostic instrument is very important for potential respondents because it provides clear directions on how to answer the questions correctly. Good instructions can help potential respondents understand the purpose of the instrument and how they should interpret the questions and provide accurate responses. This can also help reduce misunderstandings or wrong interpretations that can interfere with the validity of measurement results.

Second, in terms of content, the average value reached 3.45. This shows that the material presented on the diagnostic instrument is very relevant and comprehensive. This assessment is also included in the "Very Good" category, which shows that the content of this instrument really supports students' understanding of waves and vibrations. This cannot be separated from the quality of 4WADI, as Caleon and Subramaniam (2010) reported. They have undergone a series of structured and systematic steps in producing 4WADI, starting from Content Validation and Piloting, Construction, Administration, and Validation.

Third, in the linguistic aspect, I-4WADI obtained an average score of 3.43, which is also included in the "Very Good" category. This shows that the language used is very precise and easy for students to understand. Linguists have guaranteed equivalence between the original language version of the instrument (4-WADI) and the Indonesian version (I-4WADI) through a forward-backward translation procedure. They are responsible for ensuring that the translation reflects the original text's exact meaning and takes into account the cultural and linguistic context of the target language. Linguist experts provide recommendations and ensure the target Indonesian students understand the translated instruments well. More than that, linguists also recommend some relevant vocabulary, avoiding phrases or terms that are ambiguous or difficult for potential respondents to understand.

Finally, the graphic aspect, which is also important in diagnostic instruments, has an average value of 3.37. Like the three previous aspects, this graphic aspect is also included in the "Very Good" category. This shows that the appearance and arrangement of the instruments are very good and support students' experience in answering questions.

Thus, the readability results of the four-level diagnostic instrument on wave and vibration materials show that it is in the "Excellent" category in all evaluated aspects. This is a very positive indication that this instrument is effective in helping students understand the questions used in the diagnostic instrument. The readability test results of I-4WADI in the Indonesian cultural context are parallel to the readability test results carried out on the original version (Caleon & Subramaniam, 2010).

3. Empirical test results

a. Item reliability

The reliability analysis results of the instrument's items are displayed in a summary, as shown in Table 1. Item reliability is a measure that indicates the extent to which the items in the test are consistent in measuring individual abilities (Zmnako & Chalabi, 2019).

Table 1. Summary of I-4WADI statistics

Index of item	Score
Infit MnSq	1.01
Infit ZStd	-0.08
Outfit MnSq	0.99
Outfit ZStd	-0.26
Separation	6.46
Reliability	0.98

The item reliability value in Table 1 is 0.98, indicating that the items in the test have a very high level of reliability in measuring individual abilities (Monaco et al., 2022). This is a positive indicator that the items are reliable in measuring the desired misconceptions. This shows that the items in the test consistently measure the ability to understand individual concepts. The reliability at the item level complements the reliability of the original 4WADI, which has yet to be reported by Caleon and Subramaniam (2010).

Item separation is a measure that indicates the extent to which the items in a test can differentiate between individuals with different abilities (Cordier et al., 2018). The item separation value of 6.46 indicates that the items in the test have a very good ability to differentiate between individuals with different abilities (Adams, Tan, & Sumintono, 2021). The higher the value of the separation item, the better the items in the test are in discriminating between individuals with different abilities.

b. Item Difficulty Distribution (LVI and Wright map)

The distribution of items at each difficulty level is visualized in Figure 2. Figure 2, the Person-item map, shows the two items (Q9 and Q12) that were easiest for students. The adequacy of the number of items, the distribution hierarchy, and the existence of gaps in the item difficulty hierarchy are indicators that support the representativeness of the content (Ha, 2021). Figure 2 shows the sufficient number of items and their distribution in each difficulty hierarchy. However, there is a significant gap between items Q12 and Q9. From the person-item map in Figure 2, there is space for item improvement by adding items that are more difficult to answer to cover all person abilities.

c. Item fit

To obtain information on item suitability for the Rasch modeling, the infit and outfit MnSq values were determined. Table 2 details the infit and outfit MnSq values of each item. The infit and outfit MnSq values provide information about the item's fit to the Rasch model. In this study, the infit MnSq values ranged from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. These values indicate that the items fit the Rasch model. Values in the range of 0.5 to 1.5 are generally considered acceptable (Li et al., 2023; Palacios-Espinosa et al., 2023; Sumintono, 2015).

The outfit Zstd value, which measures standard residuals, ranges from -3.03 to 3.65. These values indicate the extent to which the observed response deviates from the expected response based on the Rasch model. The range of values indicates that some items may not fit the model because values outside the range of -2 to 2 are usually considered problematic (Rusmansyah & Almubarak, 2020).

d. Unidimensionality

Table 2 shows the unidimensionality of the items used. Residual Principal Component Analysis has been used to determine the unidimensionality of instruments (Han, 2022; Liu & Boone, 2023). This unidimensionality indicates that the instrument measures one single dimension in the latent factors investigated in this study (Scoulas, Aksu Dunya, & De Groote, 2021; Simpelaere, Van Nuffelen, De Bodt, Vanderwegen, & Hansen, 2017). For the instrument developed to be valid in measuring latent factors or unidimensionality, it must meet the standard that the score of raw variances explained by measure exceeds 30% (Chou & Wang, 2010). However, the diagnostic instrument developed had a score of raw variances explained by the measure of 28.4%. Even though it is below the 30% limit, this figure shows that the unidimensionality of the instrument being developed remains relatively good.

```
MEASURE    Person - MAP - Item
               <more>|<rare>
   1               .  +T
                   .  |
                   #  |
                   .  |  Q11
                   #  |  Q1
                   .  |S Q5
                   .  |  Q4
                   . T|
                      |  Q8
                   #  |
                      |M Q2     Q7
                #### |  Q10    Q6
   0            .## S+  Q3
              .####  |
               .###  |
              .####  |S Q9
             ######  |
            .###### M|
       #############  |
           .########  |
            #######   |T
                      |
              .### S|
               .##   |  Q12
                      |
  -1            ####  +
                      |
                     T|
                   .  |
                      |
                      |
                   .  |
                      |
                      |
                      |
                      |
                      |
  -2               .  +
               <less>|<freq>
  EACH "#" IS 4: EACH "." IS 1 TO 3
```

Figure 2. Distribution of items in the person-item map (Wright map)

Table 2. The fit of Items to Rasch Models

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
1	743	306	.49	.07	1.40	3.03	1.25	1.59	.13	.35	62.7	64.6	Q1
2	878	306	-.04	.06	1.07	.99	1.10	1.13	.46	.39	19.0	25.2	Q2
3	931	306	-.19	.05	.83	-3.02	.83	-2.36	.43	.40	26.5	22.9	Q3
4	787	306	.28	.06	1.12	1.21	1.16	1.29	.10	.37	52.3	53.4	Q4
5	771	306	.35	.07	.94	-.57	.87	-.96	.44	.36	53.6	54.4	Q5
6	885	306	-.06	.05	1.05	.80	1.03	.38	.45	.39	19.6	25.3	Q6
7	860	306	.02	.06	.76	-3.53	.74	-3.03	.54	.39	33.7	30.5	Q7
8	828	306	.12	.06	.81	-2.39	.71	-3.01	.58	.38	41.8	37.3	Q8
9	1020	306	-.42	.05	1.03	.52	.98	-.30	.43	.40	16.0	20.7	Q9
10	891	306	-.08	.05	.90	-1.52	.89	-1.30	.35	.39	21.6	25.3	Q10
11	730	306	.56	.08	.88	-.95	.96	-.21	.29	.34	65.4	68.5	Q11
12	1252	306	-1.03	.05	1.34	4.45	1.36	3.65	.23	.35	22.2	30.8	Q12
MEAN	881.3	306.0	.00	.06	1.01	-.1	.99	-.3			36.2	38.2	
P.SD	137.1	.0	.41	.01	.19	2.3	.19	1.9			17.4	16.5	

Table 3. Unidimensionality of items

Table of STANDARDIZED RESIDUAL variance in Eigenvalue units = Item information units		Eigenvalue	Observed		Expected
Total raw variance in observations	=	16.7485	100.0%		100.0%
Raw variance explained by measures	=	4.7485	28.4%		28.4%
Raw variance explained by persons	=	.5610	3.3%		3.4%
Raw Variance explained by items	=	4.1875	25.0%		25.1%
Raw unexplained variance (total)	=	12.0000	71.6%	100.0%	71.6%
Unexplned variance in 1st contrast	=	2.0762	12.4%	17.3%	
Unexplned variance in 2nd contrast	=	1.3695	8.2%	11.4%	
Unexplned variance in 3rd contrast	=	1.2407	7.4%	10.3%	
Unexplned variance in 4th contrast	=	1.1866	7.1%	9.9%	
Unexplned variance in 5th contrast	=	1.1818	7.1%	9.8%	

CONCLUSION

A four-level diagnostic instrument on wave and vibration materials has been developed through an adaptation process in the Indonesian cultural context. The results of adapting 4WADI to the Indonesian cultural context have been carried out into I-4WADI, and the results show that the four-level type diagnostic instrument on wave and vibration material is appropriate to the cultural context of high school students in Indonesia—the results of the expert assessment of content validity range from 0.85 to 1.00. The readability level is very good, with an average score of 3.37. Empirically, I-4WADI has infit MnSq values ranging from 0.81 to 1.40, and the outfit MnSq values ranged from 0.71 to 1.36. The instrument reliability value is 0.98. Therefore, I-4WADI meets the criteria for validity and reliability.

This research has the potential to make a significant contribution to curriculum development and improve the quality of education in Indonesia through the instrument adaptation process that has been carried out. However, these findings cannot be generalized to a wider context because the heterogeneity of respondents is limited to public high school students. In future research, the heterogeneity of respondents from private and vocational schools can be considered by future researchers so that it will provide a richer and more comprehensive view.

ACKNOWLEDGMENTS

We want to send our greatest gratitude to the Directorate of Research and Community Service, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for granting research funds number: 015/PFR/LPPM UAD/VI/2023 in the Fundamental Research scheme.

REFERENCES

Adams, D., Tan, M. H. J., & Sumintono, B. (2021). Students' readiness for blended learning in a leading Malaysian private higher education institution. *Interactive Technology and Smart Education*, 18(4), 515–534. Retrieved from https://doi.org/10.1108/ITSE-03-2020-0032

Assimi, E., Janati I, R., Zerhane, R., & Boubih, S. (2022). The use of a three-tier diagnostic test to investigate conceptions related to cell biology concepts among pre-service teachers of life and earth sciences. *Journal of Biological Education*, 1–28. Retrieved from https://doi.org/10.1080/00219266.2022.2134175

Aykutlu, I., Ensari, Ö., & Bayrak, C. (2023). Prospective teachers' conceptual understanding of the polarization of light. *European Journal of Physics*, 44(1), 015701. Retrieved from https://doi.org/10.1088/1361-6404/ac93cc

Bani-Salameh, H. N. (2017). How persistent are the misconceptions about force and motion held by college students? *Physics Education*, 52(1), 14003.

Brundage, M. J., & Singh, C. (2023). Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics. *Physical Review Physics Education Research*, 19(2), 020112. Retrieved from https://doi.org/10.1103/PhysRevPhysEducRes.19.020112

Caleon, I. S., & Subramaniam, R. (2010). Do Students Know What They Know and What They Don't Know? Using a Four-Tier Diagnostic Test to Assess the Nature of Students' Alternative Conceptions. *Research in Science Education*, 40(3), 313–337. Retrieved from https://doi.org/10.1007/s11165-009-9122-4

Caycho-Rodríguez, T., Vilca, L. W., Cervigni, M., Gallegos, M., Martino, P., Portillo, N., & Burgos Videla, C. (2022). Fear of COVID-19 scale: Validity, reliability and factorial invariance in Argentina's general population. *Death Studies*, *46*(3), 543–552. doi:10.1080/07481187.2020.1836071

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A Review of Studies About Four-Tier Diagnostic Tests in Physics Education. *Journal of Turkish Science Education*, 19(4), 1291–1311. Retrieved from https://doi.org/10.36681/tused.2022.175

Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23(2), 485–493. Retrieved from https://doi.org/10.1007/s11136-013-0487-5

Chou, Y.-T., & Wang, W.-C. (2010). Checking Dimensionality in Item Response Models with Principal Component Analysis on Standardized Residuals. *Educational and Psychological Measurement*, 70(5), 717–731. Retrieved from https://doi.org/10.1177/0013164410379322

Cordier, R., Speyer, R., Schindler, A., Michou, E., Heijnen, B. J., Baijens, L., & Joosten, A. V. (2018). Using Rasch Analysis to Evaluate the Reliability and Validity of the Swallowing Quality of Life Questionnaire: An Item Response Theory Approach. *Dysphagia*, 33(4), 441–456. Retrieved from https://doi.org/10.1007/s00455-017-9873-4

Dwiastuti, E., Sukarelawan, M. I., & Sriyanto, S. (2022). Learning Management System Acceptance Rate Among Vocational School Physics' Student. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 6(1), 10–18. Retrieved from https://doi.org/10.30599/jipfri.v6i1.1183

Guo, H., Lu, R., Johnson, M. S., & McCaffrey, D. F. (2022). Alternative Methods for Item Parameter Estimation: From CTT to IRT. *ETS Research Report Series*, 2022(1), 1–16. Retrieved from https://doi.org/10.1002/ets2.12355

Ha, H. T. (2021). A Rasch-based validation of the Vietnamese version of the Listening Vocabulary Levels Test. *Language Testing in Asia*, 11(1), 16. Retrieved from https://doi.org/10.1186/s40468-021-00132-7

Haidar, D. A., Yuliati, L., & Handayanto, S. K. (2020). The effect of inquiry learning with scaffolding on misconception of light material among fourth-grade students. *Jurnal Pendidikan IPA Indonesia*, 9(4), 540–553. Retrieved from https://doi.org/10.15294/jpii.v9i4.22973

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PLOS ONE*, 17(9), e0271992. Retrieved from https://doi.org/10.1371/journal.pone.0271992

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. Retrieved from https://doi.org/10.12973/tused.10269a

Larasati, P. E., Supahar, & Yunanta, D. R. A. (2020). Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material. *Journal of Physics: Conference Series*, 1440(1), 012020. Retrieved from https://doi.org/10.1088/1742-6596/1440/1/012020

Lemmer, M., Kriek, J., & Erasmus, B. (2020). Analysis of Students' Conceptions of Basic Magnetism from a Complex Systems Perspective. *Research in Science Education*, 50(2), 375–392. Retrieved from https://doi.org/10.1007/s11165-018-9693-z

Li, D., Zong, X., Huang, Q., Wu, F., Huang, Y., Ge, Y., … Yuan, C. (2023). Validation of the simplified Chinese version of PROMIS Parent Proxy-25 Profile in parents of children with cancer. *Journal of Pediatric Nursing*, in press. Retrieved from https://doi.org/10.1016/j.pedn.2023.05.016

Lim, H. L., & Poo, Y. P. (2021). Diagnostic test to assess misconceptions on photosynthesis and plant respiration: Is it valid and reliable? *Jurnal Pendidikan IPA Indonesia*, 10(2), 241–252. Retrieved from https://doi.org/10.15294/jpii.v10i2.26944

Ling Lee, W., Chinna, K., & Sumintono, B. (2020). Psychometrics assessment of HeartQoL questionnaire: A rasch analysis. *European Journal of Preventive Cardiology*, 2047487320902322. Retrieved from https://doi.org/10.1177/2047487320902322

Liu, X., & Boone, W. J. (2023). Introduction to Advances in Applications of Rasch Measurement in Science Education. In X. Liu & W. J. Boone (Eds.), *Advances in Applications of Rasch Measurement in Science Education. Contemporary Trends and Issues in Science Education* (pp. 1–18). Springer. Retrieved from https://doi.org/10.1007/978-3-031-28776-3_1

Luce, M. R., & Callanan, M. A. (2020). Family Conversations About Heat and Temperature: Implications for Children's Learning. *Frontiers in Psychology*, 11(August), 1–18. Retrieved from https://doi.org/10.3389/fpsyg.2020.01718

Maknun, J., & Marwiah, M. (2022). Remediation of Misconceptions Vocational High School Students on the Concept of Static Fluids using the Conceptual Change Model. *Journal of Technical Education and Training*, 14(2). Retrieved from https://doi.org/10.30880/jtet.2022.14.02.005

Matheny, L. M., & Clanton, T. O. (2020). Rasch Analysis of Reliability and Validity of Scores From the Foot and Ankle Ability Measure (FAAM). *Foot and Ankle International*, 41(2), 229–236. Retrieved from https://doi.org/10.1177/1071100719884554

Monaco, E., Pisanu, G., Carrozzo, A., Giuliani, A., Conteduca, J., Oliviero, M. & Ferretti, A. (2022). Translation, cross-cultural adaptation, and validation of the Italian version of the anterior cruciate ligament–return to sport after injury (ACL-RSI) scale and its integration into the K-STARTS test. *Journal of Orthopaedics and Traumatology*, 23(1), 1–8. Retrieved from https://doi.org/10.1186/s10195-021-00622-7

Muñiz, J., Elosua, P., Padilla, J.-L., & Hambleton, R. K. (2016). Test Adaptation Standards for Cross-Lingual Assessment. In C. S. Wells, M. Faulkner-Bond, & E. Hambleton (Eds.), *Educational Measurement: From Foundations to Future* (pp. 291–304). The Guilford Press.

Negoro, R. A., & Karina, V. (2019). Development of a Four-Tier Diagnostic Test for Misconception of Oscillation and Waves. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 5(2), 69–76. Retrieved from https://doi.org/10.21009/1.05201

Palacios-Espinosa, X., Pedraza, R. S., Gómez-Carvajal, A.-M., Botero-Meneses, J. S., Escallón, D. M., & Leal, D. A. (2023). Psychometric Properties of the Death Anxiety Scale for Adult Chronic Patients. *Trends in Psychiatry and Psychotherapy*, in press. Retrieved from https://doi.org/10.47626/2237-6089-2023-0630

Rusmansyah, R., & Almubarak, A. (2020). Students' Cognitive Analysis using Rasch Modeling as an Assessment for Planning of Strategies in Chemistry Learning. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 5(3), 222. Retrieved from https://doi.org/10.20961/jkpk.v5i3.40936

Scoulas, J. M., Aksu Dunya, B., & De Groote, S. L. (2021). Validating students' library experience survey using rasch model. *Library & Information Science Research*, 43(1), 101071. Retrieved from https://doi.org/10.1016/j.lisr.2021.101071

Simpelaere, I. S., Van Nuffelen, G., De Bodt, M., Vanderwegen, J., & Hansen, T. (2017). Validation of the Dutch version of the Swallowing Quality-of-Life Questionnaire (DSWAL-QoL) and the adjusted DSWAL-QoL (aDSWAL-QoL) using item analysis with the Rasch model: a pilot study. *Health and Quality of Life Outcomes*, 15(1), 66. Retrieved from https://doi.org/10.1186/s12955-017-0639-3

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian Student Misconceptions in Science Concepts. *Heliyon*, 8(9), e10720. Retrieved from https://doi.org/10.1016/j.heliyon.2022.e10720

Somroob, S., & Wattanakasiwich, P. (2017). Investigating student understanding of simple harmonic motion. *Journal of Physics: Conference Series*, 901, 012123. Retrieved from https://doi.org/10.1088/1742-6596/901/1/012123

Stefanou, M., Stylos, G., Georgopoulos, K., & Kotsis, K. (2023). Primary Preservice Teachers' Misconceptions and Reasoning of Thermal Concepts in Everyday Contexts. *The International Journal of Learning in Higher Education*, 31(1), 127–157. Retrieved from https://doi.org/10.18848/2327-7955/CGP/v31i01/127-157

Suhendi, H. Y., & Ardiansyah, R. (2021). Development of HATRADI : a four tier test for diagnostic student misconception in heat transfer concept. *Journal of Physics: Conference Series*, 1918(2). Retrieved from https://doi.org/10.1088/1742-6596/1918/2/022015

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., Soeharto, S., & Hikmah, F. N. (2021). Rasch Analysis to Evaluate the Psychometric Properties of Junior Metacognitive Awareness Inventory in the Indonesian Context. *Jurnal Pendidikan IPA Indonesia*, 10(4), 486–495. Retrieved from https://doi.org/10.15294/jpii.v10i4.27114

Sukarelawan, M. I., Jumadi, J., Kuswanto, H., & Thohir, M. A. (2021). The Indonesian Version of the Physics Metacognition Inventory: Confirmatory Factor Analysis and Rasch Model. *European Journal of Educational Research*, *10*(4), 2133–2144. doi:10.12973/eu-jer.10.4.2133

Sukarelawan, M. I., Puspitasari, A. D., Rahmatika, Z. 'Amalia, Dennis, D., Ishafit, I., Hikmah, F. N., & Sulistyo, E. N. (2022). Online Learning Using Google Incorporated for Student High School: Mapping Motivation Using Rasch Model in Physics Learning. *Berkala Ilmiah Pendidikan Fisika*, 10(2), 216–224. Retrieved from https://doi.org/10.20527/bipf.v10i2.12209

Sukarelawan, M. I., Puspitasari, A. D., Sulisworo, D., Kuswanto, H., & Jumadi. (2022). A shift in the conceptual understanding of physics students through the wright map. *Jurnal Pendidikan Dan Pengajaran*, 55(1), 127–141. Retrieved from https://doi.org/10.23887/jpp.v55i1.37155

Sukarelawan, M. I., Sriyanto, S., Puspitasari, A. D., Sulisworo, D., & Hikmah, U. N. (2021). Four-Tier Heat and Temperature Diagnostic Test (4T-HTDT) to Identify Student Misconceptions. *JIPFRI (Jurnal Inovasi Pendidikan Fisika Dan Riset Ilmiah)*, 5(1), 1–8. Retrieved from https://doi.org/10.30599/jipfri.v5i1.856

Sumintono, B. (2015). Pemodelan Rasch pada Asesmen Pendidikan: Suatu Pengantar. *Konferensi Guru Dan Dosen Nasional (KGDN) 2015*, (November), 1–14. Retrieved from http://eprints.um.edu.my/id/eprint/15603%0A

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi model rasch untuk penelitian ilmu-ilmu sosial [Rasch model application for social sciences research]*. Cimahi: Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). *Aplikasi Pemodelan Rasch pada Asesmen Pendidikan*. Cimahi: Trim Komunikata.

Syahrial, S., Ilmah, M., Yahmin, Y., Munzil, M., & Muntholib, M. (2023). Remediation of chemistry teachers' misconceptions about covalent bonding using cognitive conflict interviews: A case study. *Journal of the Serbian Chemical Society*, 88(2), 211–221. Retrieved from https://doi.org/10.2298/JSC220117073S

Yuhanna, W. L., Al Muhdhar, M. H. I., Gofur, A., & Hassan, Z. (2021). Self-reflection assessment in vertebrate zoology (Sravz) using rasch analysis. *Jurnal Pendidikan IPA Indonesia*, 10(1), 35–47. Retrieved from https://doi.org/10.15294/jpii.v10i1.25603

Zhao, C., Zhang, S., Cui, H., Hu, W., & Dai, G. (2023). Middle school students' alternative conceptions about the human blood circulatory system using four-tier multiple-choice tests. *Journal of Biological Education*, 57(1), 51–67. Retrieved from https://doi.org/10.1080/00219266.2021.1877777

Zmnako, S. S. F., & Chalabi, Y. I. (2019). Cross-cultural adaptation, reliability, and validity of the Vertigo symptom scale-short form in the central Kurdish dialect. *Health and Quality of Life Outcomes*, 17(1), 1–12. Retrieved from https://doi.org/10.1186/s12955-019-1168-z