JPPIPA 9(11) (2023)

Jurnal Penelitian Pendidikan IPA

Journal of Research in Science Education

http://jppipa.unram.ac.id/index.php/jppipa/index

The Development of Instrument of Emotional Climate and Attitude Measurement on Science Learning Environment

Hasna Rashifah[1*], Yuli Rahmawati[1], Riyadi[1]

[1] Department of Educational Research and Evaluation, Graduate Program, State University of Jakarta, Jakarta, Indonesia.

Received: August 18, 2023
Revised: September 7, 2023
Accepted: November 25, 2023
Published: November 30, 2023

Corresponding Author:
Hasna Rashifah
hasnarashifah97@gmail.com

DOI: 10.29303/jppipa.v9i11.5029

Abstract: The purpose of this research is to enrich the discussion of the component of developing instrument as a realization of student reflection on their learning environment, and as an evaluation by researchers, teachers, and principals to increase students' motivation and achievement. The approach of the research is a quantitative research with developing the instrument method. The research is conduct on 1.067 students of public middle school in Jakarta. The result demonstrate that the instrument that developed and validated from the instrument of Classroom Emotional Climate (CEC) and the instrument of Test Of Science-Related Attitudes (TOSRA) have eight dimensions are, collaboration, motivation, care, clarity, attitude to scientific inquiry, enjoyment of lessons, adoption of scientific attitudes, and sosial implications of science, used LISREL to test Confirmatory Factor Analysis are 37 items. The further analysis use Rasch Fit Item value is 36 items and 589 person, have five items level of difficulty and person abilities are also clustered into five levels. Besides that, Cronbach's alpha value is 0.92, Person reliability is 0.88 and item reliability is 0.99, shows that the instrument results is reliabel, therefore able to use.

Keywords: Attitude; Emotional climate; Learning environment; Science

Introduction

Science learning is learning that is viewed as a method to develop students' critical thinking skills to advance in scientific knowledge or technology to face the development era. This is according to Sutiani et al. (2021) stated that learning science has develop critical thinking skills as a method to enhance students' competencies in problem solving and science innovations. Besides that, science learning also have a significant role in developing students' potential to be well prepared to enter the real world (Rahmawati et al., 2020). However many students cannot relish in learn science, students assume that they are not competent of learning science and learning science is not consider essential (Ng, 2021). According to Ulumiyah et al. (2022) that students have lack interest in learning science. It can lead decrease motivation on students learning as an essential aspect in supporting student achievement (Pratama et al., 2021). According to Salta et al. (2015) that elementary school students have higher motivation than middle school students, meanwhile middle school students have higher learning motivation than high school students. Thus, it can be said that students' motivation decreases with age (Yeung & Mclnerney, 2005) due to hormonal factors, the effects of puberty, changing desire in early adolescence, and focus more on peer relationships (Wijsman et al., 2019).

The declining attitude of Indonesian students towards natural science can be seen from the results of the TIMMS study and the results of the PISA assessment which are still low at the international level. The results of these assessments can illustrate the strength of students in Indonesia. In 2018 participants from Indonesia experienced a decline, this was due to a decrease in the average PISA score caused by weak understanding of information (Kemendikbud, 2019). One of the results of the TIMMS study and the low PISA assessment results can be seen in junior high school, namely in natural science subjects. To increase students' motivation, a positive learning environment is required as an effort to support learning requirements through emphasizing student participation (Hanrahan, 1998).

How to Cite:
Rashifah, H., Rahmawati, Y., & Riyadi. (2023). The Development of Instrument of Emotional Climate and Attitude Measurement on Science Learning Environment. *Jurnal Penelitian Pendidikan IPA*, *9*(11), 9313–9323. https://doi.org/10.29303/jppipa.v9i11.5029

According to Henderson et al. (2000) that the learning environment is a students' perspective of psychosocial aspects as transparency, cohesiveness and integration in learning that will lead students' participation and reflection in the class. Students' reflections in class can include the students' opinions about their academic capabilities that contain of students' perceptions of class rigidity, interactions with teachers and classmates, and students' contribution in class, which referred as classroom emotional climate (Barr, 2016). According to Talton et al. (1987) that the learning environment, not apart from the attitudes towards knowledge that is a crucial predictor of science learning achievement. Thus, the attitude and classroom emotional climate is vital to determine science learning achievement in the classroom.

However, the attitude and emotional climate of each country's classroom is different due to the different cultural values in each country. Indonesia is a country that pays attention to cultural values, one of which is the interpersonal relationship between teachers and students which requires students to respect their elders, including teachers. This is in accordance with Agency (2007) Indonesian society is governed by Pancasila (five principles) as an ideology, implicitly regulating the interaction between the young and the old, the younger generation is expected to follow the rules of the older generation which is reflected in the hierarchical and monotonous characteristics of the school system so that culture can affect interpersonal relationships between students and teachers (Telli et al., 2007).

Based on the research background above, this research developing the instrument since the instrument is the process of selecting a scale or dimension which prominent to observe, record and measure students' behavior and learning environment accurately using the scientific method. According to Walker et al. (2005) that developing and validating instrument is to measure, observe the attitudes or behavior of participants in the research, and collect the scores on instruments to accept or reject the theories (Creswell & Creswell, 2018).

The purpose of this research is to enrich the discussion of the component of developing instrument as a realization of student reflection on their learning environment, and as an evaluation by researchers, teachers, and principals to increase students' motivation and achievement. This due to the research of learning environment for decades has established that the educational environment is a consistent determinant on students' interests, cognitive and affective (Lederman & Sandra, 2014). The instrument developed and validated in this research derived from the instrument of Classroom Emotional Climate (CEC) to testing emotional climate by Fraser et al. (2021), and for attitudes test using the instrument of Test Of Science-Related Attitudes (TOSRA) by Fraser (1981). Therefore, the development and validation of the instrument of emotional climate and attitudes measurement on science learning environment in middle schools, as an effort to increase motivation and achievement of middle school students in Indonesia.

Method

The approach of the research was a quantitative research, that identifying problem based on the circumstances on the ground or as per requirement, with numerical data from the respondents. The instrument that developed were the Classroom Emotional Climate (CEC) and the Test Of Science-Related Attitudes (TOSRA), through selecting and modifying the dimension that appropriate in cultural values in Indonesia at the middle school level, especially in Science subjects. Then, the instruments were arranged into a questionnaire items based on the Likert's scale (five categories) i.e. Always - Often - Sometimes - Seldom – Never, which were online distributed used the Google Form.

The method in the research using the development procedure of Recker et al. (2010) that have five steps are, item creation, substrate identification, item identification, item revision, instrument validation. In addition, the development of the instrument according to Damanik (2014) have six steps are, identifying and developing prominent scales, modifying the instrument, translating the instrument into indigenous language, testing the instrument, analysing data and ethical considerations. Then, in this research using a modification of the method from Recker et al. (2010) and Damanik (2014) with the following flow of instrument development:

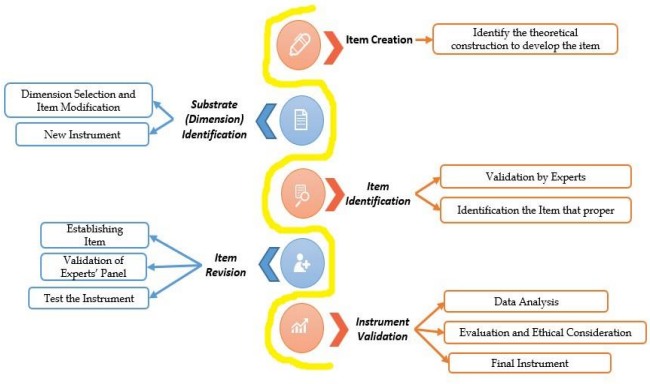

Figure 1. The Flow of instrument development

The population in the research was all the public middle school students in DKI Jakarta region. The sample selection used cluster random sampling with the

first step was cluster random sampling technique. It was found the regency or city in the East Jakarta, South Jakarta, and West Jakarta. The second step used cluster random sampling technique. Determining the location derived from the subdistrict in each regency or city of East Jakarta, South Jakarta, and West Jakarta. It found Matraman subdistrict, Kebayoran Baru subdistrict and Palmerah subdistrict. The third step used the technique of withdraw dimension response of students using the Slovin 1960's formula found 1,067 students. This technique was used to acquire the representative results.

Based on the item creation, substrate identification and item identification, the results of determining the dimensions and items of the research were eight dimensions with 64 items as follows:

Table 1. The Instrument Guidelines

Dimension	item	Total
Collaboration	1,2,3,4,5,6,7,8	8
Motivation	9,10,11,12,13,14,15,16,	8
Care	17,18,19,20,21,22,23,24	8
Clarity	25,26,27, 28,29,30,31,32	8
Attitude of Scientific Inquiry	33,34,35,36,37,38,39,40	8
Enjoyment of Science Lessons	41,42,43,44,45,46,47,48	8
Adoption of Scientific Attitudes	49,50,51,52,53,54,55,56	8
Sosial Implications of Science	57,58,59,60,61,62,63,64	8

The result of instrument revision was not apart from re-screening by experts' panel to the match of dimensions, indicators and items turn into feasible. The further step was the experts' panel validation by ten experts' panels that composed of evaluation expert, language expert, and material experts. Experts' panel pact was analysed using Fleiss Kappa analysis, with the following criteria:

Table 2. Fleiss Kappa Value (Landis & Koch, 1977)

Fleiss	Deal Power
0	Low
0.1 - 0.20	Weak
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	High
0.81 - 1.00	Almost Perfect

Afterward, the data analysis was conduct through Confirmatory Factor Analysis (CFA) with the loading factor criteria. According to Hair et al. (2018) that if the sample used more than 350 respondents, then the loading factor criteria with the sample of 1.067 was 0.30 used in the research. Then, the Construct Reliability (CR) and the Average Variance Extracted (AVE) with criteria (CR) ≥ 0.70 (JR et al., 2010), criteria (AVE) 0.50 < AVE < 1.00 (dos Santos & Cirillo, 2023). Afterward, the Model Fit Index or model feasibility with criteria according to Hair et al. (2018) in table 3 as follow:

Table 3. Model Fit Index

Model Fit Index	Cut-Off Value
Chi-Squere	≤ 2df
RMSEA	≤0.08
GFI	≥0.90
AGFI	≥0.90
TLI	≥0.95
CFI	≥0.95

Further analysis used Rasch model that can be seen from unidimensional, monotonization, and fit item. In unidimensionality that demonstrated value of *raw variance explain by measures* and value of *raw unexplained variance* with a criteria of *raw variance explain by measures*, according to Sulsilah et al. (2023) as follow:

Table 4. Value of Raw Variance Explain by Measures

Statistic	Fit Index	Interpretation
Raw Variance explained by measure	>20%	Accepted
	>40%	Good
	>60%	Extremely Good

The purpose of monotonization (rating scale) in the test was to verify the rating scale to say whether the item was confused or not for respondents. According to Sumintono & Widhiarso (2015) that the rating scale can be seen from the level of increase in logit on the Observed Average that intend to ensure that respondents able to distinguish between answer choices, besides that an increase in the Andrich Threshold was to test whether the polytomies value used was correct or not and the increase in the Category measure was to verify the not confused item the respondents..

Afterward, screening of item and person was conduct to ensure whether the item and person match or not. According to Bond & Fox (2015), from the Outfit Mean Square (MNSQ), Outfit Z Standard Estimate (ZSTD) and Point Measure Correlation values with the following criteria:

Table 5. Fit Index of MNSQ, ZSTD, and Point Measure Correlation

Statistic	Fit Index
Outfit Mean Swuare (MNSQ)	0.50-1.50
Outfit Z Standard Estimate (ZSTD)	-2.00-2.00
Point Measure correlation	0.40-0.85

After the item and person were fit (feasible), then the reliability analysis of the Rasch Analysis used winstep to provide all information about the quality of the instrument. The following reliability criteria in the Rasch Analysis:

Table 6. Reliability in the Rasch Analysis (Fisher, 2007)

Statistic	Fit index	Interpretation
Item and person reliability	<0.67	Very Low
	0.67-0.80	Low
	0.81-0.90	Moderate
	0.91-0.94	Good
	>0.94	Very Good
Item and person separation	<2	Very Low
	2-3	Low
	3-4	Moderate
	4-5	Good
	>5	Very Good

Result and Discussion

Analysis of experts' panel for the pact used SPSS version 26, i.e. Fleiss Kappa analysis. This expert validation aims to review, provide input, assessment (Siagian et al., 2023) so that researchers make improvements to the instrument (Blegur et al., 2023). As well as considering reasonable expert opinion to identify weaknesses and strengths of the instrument (Alarcon et al., 2017; Fernández-Gómez et al., 2020).

This study uses fleiss Kappa agreement which is an analytic statistic that allows to assess the level of agreement among three or more raters who independently rate a set of items using an instrument with a certain number (Fernández-Gómez et al., 2020). The result showed that the Fleiss Kappa pact had value 0.703, means it had a high agreement (Landis & Koch, 1977). Similarly to Jonsdottir et al. (2023) stated that value of 0.66 up to 0.80 had a percentage of agreement around 71-100%. Then, the instrument was tested to 1.067 respondents.

Confirmatory Factor Analysis (CFA)

The result in table 7 used LISREL, the instrument had eight dimensions were, collaboration (KBR), motivation (MTV), care (PDL), clarity (KJL), attitude of scientific inquiry (SPI), enjoyment of science lessons (KDP), adoption of scientific attitudes (ASI), and social implication of science (ISS).

The result showed that Confirmatory Factor Analysis (CFA) test with eight dimensions and had 64 items turn into 40 items was valid with Standardized Loading Factor (SLF) ≤ 0.30. The Confirmatory Factor Analysis (CFA) in the research was to verified total of dimension in the instrument and the relationship between items. According to Hoyle (2023) state that CFA intend to provide a high comprehend on the quantity and characteristic in covariant among the indicators.

Table 7. Standardized Loading Factor

Dimension	Item	Standardized Loading Factor	Note
KBR	N4	0.48	Valid
	N7	0.42	Valid
	N8	0.62	Valid
MTV	N10	0.34	Valid
	N11	0.66	Valid
	N12	0.40	Valid
	N13	0.40	Valid
	N14	0.37	Valid
	N15	0.65	Valid
	N16	0.76	Valid
PDL	N19	0.49	Valid
	N20	0.59	Valid
	N21	0.35	Valid
	N22	0.35	Valid
KJL	N23	0.64	Valid
	N24	0.73	Valid
	N27	0.51	Valid
	N28	0.57	Valid
	N31	0.58	Valid
	N32	0.63	Valid
SPI	N36	0.59	Valid
	N39	0.71	Valid
	N40	0.73	Valid
	N41	0.53	Valid
	N42	0.52	Valid
KDP	N43	0.72	Valid
	N44	0.43	Valid
	N45	0.60	Valid
	N46	0.43	Valid
	N47	0.71	Valid
	N48	0.42	Valid
	N49	0.30	Valid
ASI	N51	0.60	Valid
	N52	0.54	Valid
	N55	0.71	Valid
	N56	0.63	Valid
	N59	0.54	Valid
ISS	N60	0.57	Valid
	N63	0.75	Valid
	N64	0.52	Valid

Afterward, using Microsoft Excel to tested Construct Reliability (CR) and Average Variance Extracted (AVE) used in this study as a reliability coefficient to intensify the reliability of questionnaire-based instruments comparable to Cronbach's alpha (Rosli et al., 2021), to acquire the reliability in the instrument and each dimension as follow.

Table 8 showed that Construct Reliability (CR) had value 0.95 ≥ 0.70, thus be able to fulfill the criteria with the high reliability category. In addition, the Construct Reliability (CR) in the table had a high category on dimension motivation was 0.72, care was 0.70, attitude of scientific inquiry was 0.72, enjoyment of science lessons was 0.77, and adoption of scientific attitude was 0.70. Besides that, there were three dimensions had a

moderate category on dimension collaboration was 0.51, clarity was 0.66, and social implications of science was 0.69. Meanwhile, the Average Variance Extracted (AVE) had value < 0.50 in the instrument and each dimension that showed less than 50% construction variance was different than other construction (Vinzi et al., 2010). According to (JR et al., 2010), AVE value < 0.50 showed the appropriate convergence. Similarly to Fornell et al. (1981) that the validity of construct convergence had already been appropriate, although the value more than 50% was due to errors.

Table 8. The Test of Reliability

Dimension	CR	AVE	Category
KBR	0.51	0.26	Moderate
MTV	0.72	0.29	High
PDL	0.70	0.30	High
KJL	0.66	0.33	Moderate
SPI	0.72	0.46	High
KDP	0.77	0.31	High
ASI	0.70	0.33	High
ISS	0.69	0.36	Moderate
Instrument	0.95	0.32	High

Afterward, the CFA test was conducted to assess the suitability and reliability in each dimension. The results found in the first was collaboration dimension that had three items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.51 and Average Variance Extracted (AVE) 0.26 that the reliability value was relative appropriate. Second was motivation dimension had seven items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.72 and Average Variance Extracted (AVE) 0.26 which had a good reliability value. Third was care dimension had six items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.71 and Average Variance Extracted (AVE) 0.30 which had a good reliability value.

Fourth was clarity dimension had four items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.66 and Average Variance Extracted (AVE) 0.33, means that the reliability value was relative appropriate. Fifth was attitude of scientific inquiry dimension had three items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.72 and Average Variance Extracted (AVE) 0.46 that had a good reliability value. Sixth was enjoyment of science lessons dimension had eight items, with two items had a loading factor < 0.30, therefore the items must be dropped out were, item 44 and item 48, then the CFA was tested with six items that produced a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.71 and Average Variance Extracted (AVE) 0.40 that had a good reliability value.

Seventh was adoption of scientific attitude dimension had five items with one item had a loading factor < 0.30, therefore the item must be dropped out was item 49, then the CFA was tested with four items that produced a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.72 and Average Variance Extracted (AVE) 0.39 that had a good reliability value. Eighth was social implications of science dimension had four items with a loading factor > 0.30, means that there were no invalid items and had Construct Reliability (CR) 0.70 and Average Variance Extracted (AVE) 0.38 that had a good reliability value. Therefore, the results of the Confirmatory Factor Analysis (CFA) test on each dimension produce 37 valid and reliable items. Afterward, the model fit index or model probability test as follows:

Table 9. Model FIt

Dimension	Chi-Square < 2df	RMSEA ≤ 0.08	GFI ≥ 0.90	AGFI ≥ 0.90	TLI ≥ 0.90	CFI ≥ 0.90
KBR	0.00	0.00	0.00	0.00	0.00	0.00
MTV	220.05	0.11	0.94	0.89	0.85	0.90
PDL	117.73	0.10	0.96	0.92	0.87	0.92
KJL	22.61	0.09	0.99	0.95	0.91	0.97
SPI	0.00	0.00	0.00	0.00	0.00	0.00
KDP	712.43	0.18	0.86	0.74	0.76	0.83
ASI	38.54	0.07	0.99	0.96	0.94	0.97
ISS	54.07	0.15	0.98	0.88	0.82	0.94
Instrument	7351.49	0.09	0.74	0.70	0.93	0.94

Based on the table 9, found chi-square value was 7351.49 > 2df, RMSEA was 0.094 ≥ 0.08, GFI was 0.74 ≤ 0.90 and AGFI was 0.70 ≤ 0.90 had not fit of the model. However, TLI value was 0.93 ≥ 0.90 and CFI value was 0.94 ≥ 0.90 had fit of the model, therefore the test could be fulfilled. In line with Widarjono (2010), the model that state to be feasible if one of the model fit test method is fill. The model fit was presented in the research to provide information for these statistical results are influenced through a large sample size (Antonietti et al.,

2023). The data supports the hypothetical model based on the eight dimensions that have been developed.

Analisis Rasch

Rasch analysis is the further analysis from the developing instrument with prerequisite analysis test Rasch modelling that consist of unidimensional, monotonization, and item fit (DiStefano & Morgan, 2010). The result of unidimensional can be seen in table 10.

Table 10. Unidimensional

Variance	Unexplned variance	Raw variance explain by measures
1	6.00%	
2	4.60%	
3	3.80%	36.60%
4	3.00%	
5	2.90%	

The results from table 10 was the raw variance explained by measures had a value 36.6%, means that it was acceptable to fulfill unidimensional requirements (Sulsilah et al., 2023). The unexplained variance value were 6.0%, 4.6%, 3.8%, 3.0% and 2.9%. Unidimension items in the instrument at the variance 1st was good, at the variance 2st, 3st, and 4st, were very good, the variance 5st was excellent, means it able to fulfill the Rasch modelling ideal (Fisher, 2007). Thus, the empirical instrument was unidimensional and was able to established construct validity. It can be seen from each indicator that it can only be explained by one underlying construct (Hair et al., 2018).

Monotonization is the assumption of monotone demonstrating that the probability of maintenance an item had increase along with improving the quality level (Giguère et al., 2023). The scale of frequency from five categories of response: Always (5), Often (4), Sometimes (3), Seldom (2), and Never (1), in the table 11 as follow:

Table 11. Scale of Monotone

Category	Observed Average	Andrich Threshold
1	0.44	NONE
2	0.51	-1.03
3	0.88	-1.01
4	1.37	1.09
5	2.23	0.95

The result of rating scale analysis demonstrated that there was an increased value on observed average from negative to positive (Andrich, 2011). Logit value from 0. 44 for scale 1 (never) to 2.23 for scale 5 (always). Increasing logit value monotonically means the respondents able to distinguished the answer choices and selecting the answer based on the scale of five categories.

Besides the Observed Average, the monotonization also observed from Andrich Threshold value that moved up from none value, to negative value, and then moved to the positive value sequentially. The increase must be a change of at least one logit from one category to another (Leyva, 2023). The result from Andrich Threshold none value move to -1.03, -1.01 and move to the positive value was 1.09, 0.95, means that the instrument option was valid, however the results in the table was not in sequence that means the instrument option should be simplified.

Afterward, testing fit item from 37 items with eight dimensions on the instrument of emotional climate and attitude measurement on Science learning environment. The result of the statistic item found an item of N7 on collaboration dimension that had Outfit MNSQ criteria value was 1.69 > 1.5, means there was not fulfilled productive criteria to measuring and ZSTD value was 7.4 > 2.00, means the data was not fit the model. From the three requisites, two requisites were MNSQ and ZSTD had not fulfilled the criteria, therefore there were no revision, then an item N7 must be dropped out. The item fit that arranged turn into 36 items from eight dimensions. This was the result of item N7 removal:

Table 12. Statistic Item

Item	Outfit		Pt. Measure Corr
	MNSQ	ZSTD	
64	1.35	3.00	0.36
59	1.12	1.20	0.43
27	1.25	2.70	0.42
8	1.31	3.70	0.45
12	1.29	4.60	0.43
10	1.22	3.40	0.44
4	1.20	3.10	0.40
56	1.06	0.60	0.44
36	1.16	2.10	0.44
28	1.08	1.00	0.43
51	1.04	0.50	0.46
52	1.15	2.20	0.50
60	1.07	0.80	0.44
32	1.04	0.60	0.51
13	1.05	0.80	0.49
14	1.03	0.50	0.45
22	1.02	0.30	0.41
20	0.78	-1.70	0.41
24	0.90	-1.10	0.52
42	0.99	-0.10	0.51
15	0.98	-0.30	0.57
40	0.92	-0.90	0.52
39	0.92	-0.80	0.45
63	0.53	-3.60	0.46
19	0.98	-0.40	0.45
55	0.85	-1.50	0.53
23	0.69	-2.70	0.47

Item	Outfit		Pt. Measure Corr
	MNSQ	ZSTD	
21	0.85	-1.50	0.43
11	0.83	-2.30	0.57
41	0.89	-1.50	0.53
45	0.87	-2.10	0.59
31	0.86	-2.50	0.55
47	0.70	-4.10	0.60
43	0.79	-3.70	0.61
46	0.76	-2.40	0.46
16	0.62	-6.10	0.66

Based on the table 12, statistic item MisFit Order Instrument able to demonstrated from Outfit Mean Square (MNSQ) 0.5 – 1.5, Outfit Z-Standards (ZSTD) -2.00 – 2.00 and Point Measure Correlation (Pt Mean Corr) 0.40 – 0.85. However, the item with ZSTD value was -2.00 – 2.00, it comes as the respondents that used > 500. This was according to Idris et al. (2021) state that ZSTD value is significantly affected through the sample. If the sample is huge (N > 500), then ZSTD value more than 3 continuously, therefore the experts were not suggest to not using ZSTD. The result of item analysis was 36 had MNSQ 0.5 - 1.5 and Pt Mean Corr. 40 - 0.85, then the item was fit and appropriate therefore the item works well. Afterward, the data of fit also demonstrated from measure relative to item difficulty as follow:

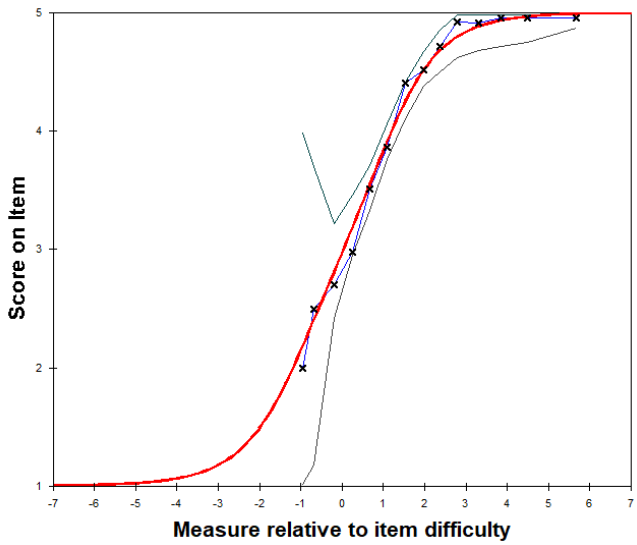

Figure 2. Item fit N15

The data of fit which appropriate to Rasch model also able to seen from Expected score ICC chart through comparing the ideal model and the data. The chart above demonstrated one of the items with a red line indicating the ideal Rasch model, meanwhile the blue line and dot (x) indicating the data obtained. The chart on N15 demonstrate that the data with the blue line was not much different from the ideal Rasch data model, which means that the data were appropriate to the Rasch ideal model.

Analysis of difficulty item also able to seen from the logit item value for item N12 on the motivation dimension with the statement "My class atmosphere is not conducive when science learning" had a logit value +1.27 indicating the item was very hard for respondents, meanwhile N63 on the implications of social science dimension with the item statement "Science practicum is only wasting time" had a logit value -1.38 was the easiest item. Analysis of item statements on item measure contains information of the mean value and standard deviation (SD), therefore the item levels could be collected. The mean value was 0.0 logit and the SD was 0.62 logit on the scale items of five response categories. Mean value 0.0 logit + SD 0.62 logit = + 0.62 logit (limit group of items with hard level); item difficulty level > 1.24 logit (limit group of item with very hard level); then if mean was 0.0 logit - SD 0.62 logit = - 0.62 logit (limit group of item with moderate level); item difficulty level -0.62 logit to -1.24 logit (limit group of items with easy level); item difficulty level < 1.24 logit (limit group of items with very easy level). Here's the item difficulty map.

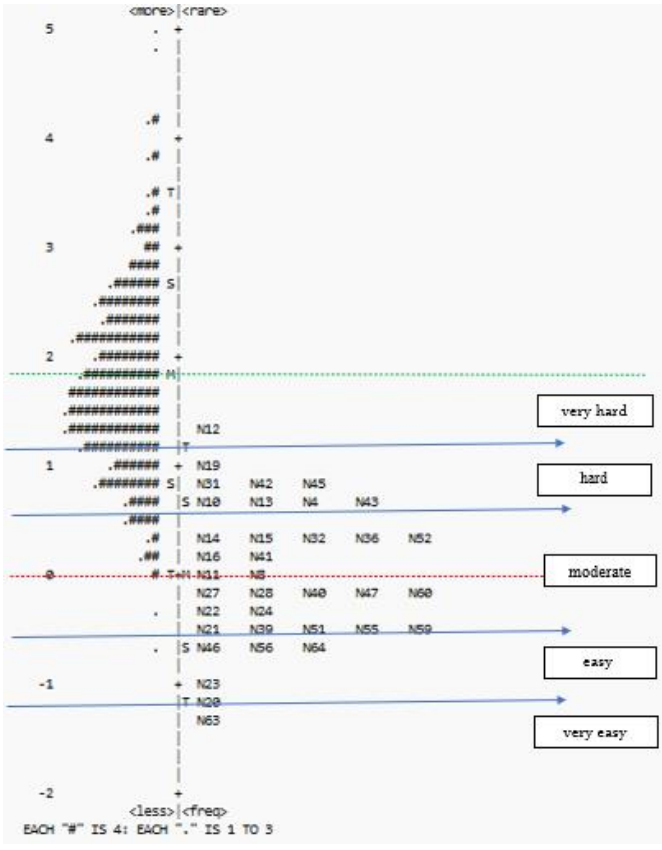

Figure 3. Wright maps

Afterward, Person Fit was conducted to fulfill the requirement for Rasch analysis, as a reliable student

response based on the student's ability level, as well as to meet the standard of reasonableness in responding to statements in the instrument (Bond & Fox, 2015; Smith, 1986), refers to the appropriate Person from the Outfit Mean Square (MNSQ) 0.5-1.5, and Point Measure Correlation 0.40 - 0.85, however the Outfit Z Standard Estimate (ZSTD) was not used since the respondents > 500 students. The results from 1,067 person was acquired 589 person fit. Then the person analysis conducted to acquire the data on respondents with high ability and low ability in answering items. It able to be seen that the higher logit values in the Person measure represent better "ability", meanwhile the higher item represent items that the more difficult to support (Balparda et al., 2020). The result demonstrated that respondents with the lowest ability were 791 respondents, and respondents with the highest ability were 410 respondents. Collecting mean value was 1.70 logit and SD was 0.86 logit. Classification of students' ability which the mean 1.70 logit + SD 0.86 logit = +2.56 logit (student response level with moderate level); from +2.56 to +3.42 logit (student response level with high level); score level > +3.42 logit (student response level with very high level); whereas if mean value was 1.70 logit – SD 0.86 logit = +0.84 logit (student response level with moderate level); from +0.84 logit to -0.02 logit (student response level with low level); score level < -0.02 (student response level with very low level).

The validity of the instrument from 589 respondents with 36 items able to be seen in the table of summary statistic from Winstep version 4.0.1 for Rasch analysis to provide information of the index of reliability instrument as an illustration of the instrument quality as follows:

Table 13. Summary Statistic

Statistic	Result
Person reliability	0.88
Item reliability	0.99
Person Separation	2.76
Item Separation	10.24
alpha Cronbach	0.92
Standard Error	0.05–0.09
P.SD	0.62
S.SD	0.63

Person reliability and item reliability, the index of the person reliability value was 0.88 and the index of item reliability value was 0.99, means the consistency of the answers from respondents was good with better quality of the item reliability. In addition, the Cronbach's alpha value reflects the interaction between items and persons with value more than 0.90 (Muslihin et al., 2022). The Cronbach's alpha resulted was 0.92, means that the interaction between persons and items was excellent.

Person and item separation index, is the estimation of the instrument to distinguish among student abilities. The results were the person separation value was 2.76 and the separation item value was 10.24 which provide information that the student separation criteria support a reliable instrument. In line with Boone et al. (2017) that value of index separation range from 0 to unlimited, and the higher values indicate better separation.

Precision of measurement, is the level of reliability of the instrument to draw conclusions. According to (Perera et al., 2018) the standard error in a good instrument is less than 0.5. The result demonstrated that range of the standard error was 0.05-0.09, thus indicating reliable and good accuracy in measurement.

Item calibration, IRT modelling process to see the balance of the scale on the item. Items are sorted and scaled allowing to the level of difficulty. The more difficult the log value above zero, the items log zero in an average difficulty level, and the easier the log value below zero, the items log zero in an easy level (Perera et al., 2018). The result demonstrated that the person standard deviation (P.SD) was 0.62 logit and the statistics standard deviation (S.SD) was 0.63 logit, the item scale ranges from -1.38 to 1.27. Then the item calibration within the range of two SD, with the sum of the two SDs were 0.62 logit + 0.63 logit = 1.25 logit. Thus the instrument indicated that there were no item misfit in the instrument.

Additional thing that requires to be understood from the output, is the test information function on the scale of five response categories. The following picture of the information function in this research.

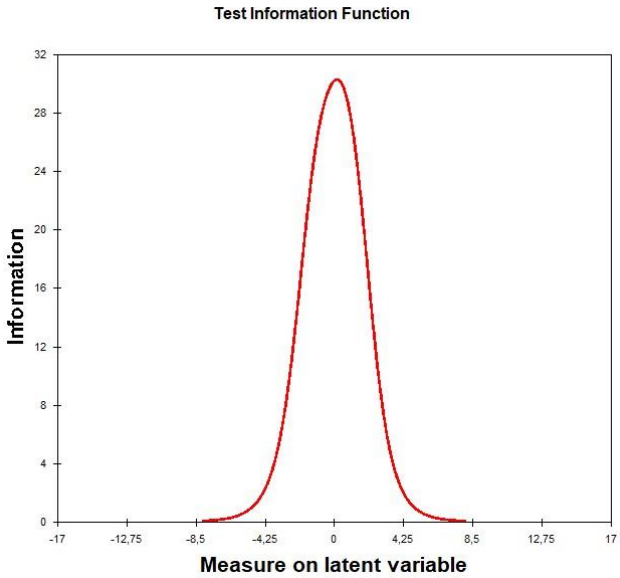

Figure 4. Information curve

The information function shows the information of measurement that found from the instrument.

According to Embretson et al. (2000) the information curve should be the form of a "plateau" is obtain a higher amount of information on a wider range. The X-axis showed the level of ability to answer statements on the instrument, and the Y-axis showed the value of the information function (Adi et al., 2022). Based on the information curve above, the function obtained was relatively high, therefore the control of the items on the instrument providing a lot of information.

Conclusion

Based on the results of the research the development of instrument of emotional climate and attitude measurement on science learning environment, found Confirmatory Factor Analysis (CFA) 37 items are valid the criteria in reliability category. Furthermore, the Rasch analysis shows the instrument is valid. The results from 36 items have five items level of difficulty are, very hard, hard, medium, easy and very easy. Person abilities are also clustered into five levels are very high, high, moderate, low and very low. Besides that, shows that the reliability instrument results is reliability. In addition, this is shows that the instrument is acceptable.

Acknowledgments
We would like to appreciate the lecturers at State University of Jakarta, SMPN 97 Jakarta, SMPN 240 Jakarta, and SMPN 101 Jakarta and also the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia who have assisted in this research.

Author Contributions
HR conducted instrument analysis, data collection, data analysis and data interpretation. YR participated in the alignment of dimensions and items developed, instrument development design and data interpretation. R participated in data analysis and data interpretation. All authors contributed to the drafting of the manuscript and approved the final manuscript.

Funding
Funding was provided by the Ministry of Education, Culture, Research, and Technology, the Republic of Indonesia, grant number: 138/E5/PG.02.00.PL/2023.

Conflicts of Interest
No conflicts of interest

References

Adi, N. R. M., Amaruddin, H., Adi, H. M. M., & A`yun, L. Q. (2022). Validity and Reliability Analysis Using the Rasch Model to Measure the Quality of Mathematics Test Items of Vocational High Schools. *Journal of Educational Research and Evaluation*, *11*(2), 103–113. Retrieved from http://journal.unnes.ac.id/sju/index.php/jere

Agency, C. I. (2007). *The world fact book: Indonesia*. The World Fact Book: Indonesia. Retrieved from https://www.cia.gov/cia/publications/factbook/print/id.html.

Alarcon, L. A. G., Trapaga, J. A. B., & Navarro, R. E. (2017). A photovoice study of older adults' conceptualizations of risk. *Journal of Aging Studies*, *26*(4), 448–458. Retrieved from http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L52096536%0A http://dx.doi.org/10.1016/j.jaging.2012.06.004

Andrich, D. (2011). *Rating scales and Rasch measurement*.

Antonietti, C., Schmitz, M. L., Consoli, T., Cattaneo, A., Gonon, P., & Petko, D. (2023). "Development and validation of the ICAP Technology Scale to measure how teachers integrate technology into learning activities." *Computers and Education*, *192*(August 2022), 104648. https://doi.org/10.1016/j.compedu.2022.104648

Balparda, K., Herrera-Chalarca, T., Silva-Quintero, L. A., Torres-Soto, S. A., & Vanegas-Ramírez, C. M. (2020). Development and validation of the "Keratoconus end-points assessment questionnaire" (KEPAQ), a disease-specific instrument for evaluating subjective emotional distress and visual function through rasch analysis. *Clinical Ophthalmology*, *14*, 1287–1296. https://doi.org/10.2147/OPTH.S254370

Barr, J. J. (2016). Developing a positive classroom climate. *The IDEA Center*, *61*(October), 1–9. Retrieved from http://files.eric.ed.gov/fulltext/ED573643.pdf

Blegur, J., Rajagukguk, C. P. M., Sjioen, A. E., & Souisa, M. (2023). Innovation of Analytical Thinking Skills Instrument for Throwing and Catching Game Activities for Elementary School Students. *International Journal of Instruction*, *16*(1), 723–740. https://doi.org/10.29333/iji.2023.16140a

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Third Edition 3rd Edition. In *Maternal & Child Nutrition* (Vol. 11, Issue supplement S2). Routledge.

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, *4*(1), 1–13. https://doi.org/10.1080/2331186X.2017.1416898

Creswell, W. J., & Creswell, J. D. (2018). Research Design: Qualitative, Quantitative adn Mixed Methods Approaches. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9). SAGE. Retrieved from http://www.ceil-conicet.gov.ar/wp-content/uploads/2015/10/Creswell-Cap-10.pdf

Damanik, E. (2014). *Science and Mathematics Education Centre Principal Leadership Style and Its Impact on School Climate and Teacher Self-Efficacy in Indonesian Schools Enceria Damanik* (Issue February). Curtin University of Technology. Retrieved from https://espace.curtin.edu.au/bitstream/handle/20.500.11937/391/200089_Damanik 2014.pdf?sequence=2

DiStefano, C., & Morgan, G. B. (2010). Evaluation of the BESS TRS-CA Using the Rasch Rating Scale Model. *School Psychology Quarterly*, *25*(4), 202–212. https://doi.org/10.1037/a0021509

dos Santos, P. M., & Cirillo, M. Â. (2023). Construction of the average variance extracted index for construct validation in structural equation models with adaptive regressions. *Communications in Statistics: Simulation and Computation*, *52*(4), 1639–1650. https://doi.org/10.1080/03610918.2021.1888122

Embretson, S. E., & Reise, S. P. (2000). Item Response Theory For Psychologists. In *Item Response Theory*. Lawrence Erlbaum Associates, Inc. https://doi.org/10.4324/9781410605269-13

Fernández-Gómez, E., Martín-Salvador, A., Luque-Vara, T., Sánchez-Ojeda, M. A., Navarro-Prado, S., & Enrique-Mirón, C. (2020). Content validation through expert judgement of an instrument on the nutritional knowledge, beliefs, and habits of pregnant women. *Nutrients*, *12*(4). https://doi.org/10.3390/nu12041136

Fisher, W. P. (2007). Rating scale instrument quality criteria. *Rasch Measurement Transactions*, *21*(1), 1095.

Fornell, C., & Larcker F., D. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *XVIII*(February), 39–50. https://doi.org/10.1177/002224378101800313

Fraser, B. J. (1981). *Test of science-related attitudes*.

Fraser, B. J., McLure, F. I., & Koul, R. B. (2021). Assessing Classroom Emotional Climate in STEM classrooms: developing and validating a questionnaire. *Learning Environments Research*, *24*(1). https://doi.org/10.1007/s10984-020-09316-z

Giguère, G., Brouillette-Alarie, S., & Bourassa, C. (2023). A Look at the Difficulty and Predictive Validity of LS/CMI Items With Rasch Modeling. *Criminal Justice and Behavior*, *50*(1), 118–138. https://doi.org/10.1177/00938548221131956

Hair.Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis*. Cengage. https://doi.org/10.1002/9781119409137.ch4

Hanrahan, M. (1998). The effect of learning environment factors on students' motivation and learning. *International Journal of Science Education*, *20*(6), 737–753. https://doi.org/10.1080/0950069980200609

Henderson, D., Fisher, D., & Fraser, B. (2000). Interpersonal behavior, laboratory learning environments, and student outcomes in senior biology classes. *Journal of Research in Science Teaching*, *37*(1), 26–43. https://doi.org/10.1002/(SICI)1098-2736(200001)37:1<26::AID-TEA3>3.0.CO;2-I

Hoyle, R. H. (2023). *Handbook of Structural Equation Modeling* (Second edi). The Guilford Publications.

Idris, M., Izmaimusah, D., & Ahmad. (2021). Instrument Development of Mathematics Learning Outcomes By the Rasch Model in Elementary School To Support the Implementation of the 2013 Curriculum. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, *7*(2), 88–102. https://doi.org/10.21009/jisae.v7i2.21830

Jonsdottir, G., Haraldsdottir, E., Sigurdardottir, V., Thoroddsen, A., Vilhjalmsson, R., Tryggvadottir, G. B., & Jonsdottir, H. (2023). Developing and testing inter-rater reliability of a data collection tool for patient health records on end-of-life care of neurological patients in an acute hospital ward. *Nursing Open*, *July 2021*, 1–9. https://doi.org/10.1002/nop2.1789

JR, H. J. F., Black, W. C., J.Babin, B., & Anderson, R. E. (2010). *Multivariate Data Analysis* (p. 161).

Kemendikbud. (2019). *Hasil PISA Indonesia 2018: Akses Makin Meluas, Saatnya Tingkatkan Kualitas*. Siaran Pers. Nomor: 397/Sipres/A5.3/XII/2019. Fraser, B. J. (1981). Test of science-related attitudes. https://www.kemdikbud.go.id/main/blog/2019/12/hasil-pisa-indonesia-2018-akses-makin-meluas-saatnya-tingkatkan-kualitas

Landis, J. R., & Koch, G. G. (1977). Landis amd Koch1977_agreement of categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lederman, N. G., & Sandra, A. K. (2014). Handbook of Research on Science Education. In *Handbook of Research on Science Education* (Vol. 2). Routledge. https://doi.org/10.1097/00003465-199001000-00018

Leyva, E. M. (2023). *A Review of the Kentucky Inventory of Mindfulness Skills under Rasch Modeling for Unidimensionality Based on Difficulty*. *Research Square,* 1–18. https://doi.org/10.21203/rs.3.rs-2034570/v1

Muslihin, H. Y., Suryana, D., Ahman, Suherman, U., & Dahlan, T. H. (2022). Analysis of the Reliability and Validity of the Self-Determination Questionnaire Using Rasch Model. *International Journal of*

Instruction, *15*(2), 207–222. https://doi.org/10.29333/iji.2022.15212a

Ng, C. (2021). What kind of students persist in science learning in the face of academic challenges? *Journal of Research in Science Teaching*, *58*(2), 195–224. https://doi.org/10.1002/tea.21652

Perera, C. J., Sumintono, B., & Jiang, N. (2018). The Psychometric Validation Of The Principal Practices Questionnaire Based On Item Response Theory. *International Online Journal of Educational Leadership*, *2*(1), 21–38. https://doi.org/10.22452/iojel.vol2no1.3

Pratama, H., Maduretno, T. W., & Yusro, A. C. (2021). Online Learning Solution: Ice Breaking Application to Increase Student Motivation. *Journal of Educational Science and Technology (EST)*, *7*(1), 117–125. https://doi.org/10.26858/est.v7i1.19289

Rahmawati, Y., Ridwan, A., Faustine, S., Syarah, S., Ibrahim, I., & Mawarni, P. C. (2020). Pengembangan Literasi Sains Dan Identitas Budaya Siswa Melalui Pendekatan Etno-Pedagogi Dalam Pembelajaran Sains. *Edusains*, *12*(1), 54–63. https://doi.org/10.15408/es.v12i1.12428

Recker, J., & Rosemann, M. (2010). A Measurement Instrument for Process Modeling Research. *Scandinavian Journal of Information Systems*, *22*(2), 3–30. Retrieved from https://eprints.qut.edu.au/39834/1/c39834.pdf

Rosli, M. S., Saleh, N. S., Alshammari, S. H., Ibrahim, M. M., Atan, A. S., & Atan, N. A. (2021). Improving Questionnaire Reliability using Construct Reliability for Researches in Educational Technology. *International Journal of Interactive Mobile Technologies*, *15*(4), 109–116. https://doi.org/10.3991/IJIM.V15I04.20199

Salta, K., & Koulougliotis, D. (2015). Research and Practice Assessing motivation to learn chemistry : adaptation and validation of Science Motivation Questionnaire II with Greek secondary school students. *Chemistry Education Research and Practice*. https://doi.org/10.1039/C4RP00196F

Siagian, A. F., Ibrahim, M., & Supardi, Z. A. I. (2023). Creative-scientific decision-making skills learning model for training creative thinking skills and student decision making skills. *Nurture*, *17*(1), 10–17. https://doi.org/10.55951/nurture.v17i1.141

Smith, R. M. (1986). Person Fit In The Rasch Model. *Hispanic Journal of Behavioral Sciences*, *9*(2), 183–205. https://doi.org/10.1177/001316448604600210

Sulsilah, H., Hidayat, A., & Ramalis, T. R. (2023). Analysis of Computational Thinking Instrument for High School Student Using Rasch Model. *Jurnal Penelitian Pendidikan IPA, 9*(3), 1445-1450. https://doi.org/10.29303/jppipa.v9i3.2771

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan RASCH Pada Assessment Pendidikan. In *AplikAsi RascH pemodelAn Pada Assessment Pendidikan* (Issue October). Trim Komunikata.

Sutiani, A., Situmorang, M., & Silalahi, A. (2021). Implementation of an Inquiry Learning Model with Science Literacy to Improve Student Critical Thinking Skills. *International Journal of Instruction*, *14*(2), 117–138. https://doi.org/10.29333/iji.2021.1428a

Talton, E. L., & Simpson, R. D. (1987). Relationships of attitude toward classroom environment with attitude toward and achievement in science among tenth grade biology students. *Journal of Research in Science Teaching*, *24*(6), 507–525. https://doi.org/10.1002/tea.3660240602

Telli, S., Den Brok, P., & Cakiroglu, J. (2007). Students' perceptions of science teachers' interpersonal behaviour in secondary schools: Development of a Turkish version of the Questionnaire on Teacher Interaction. *Learning Environments Research*, *10*(2), 115–129. https://doi.org/10.1007/s10984-007-9023-2

Ulumiyah, D., Sumantri, M. S., Rahmawati, Y., & Iasha, V. (2022). An Analysis of Science Literacy Ability Elementary School Students. *Jurnal Basicedu*, *6*(3), 3544–3553. https://doi.org/10.31004/basicedu.v6i3.2623

Vinzi, V. E., Chin, W. M., Hanseler, J., & Wang, H. (2010). Handbook of Partial Least Squares: Concepts, Methods and Applications. In *Springer Handbooks of Computational Statistics*. Springer. Retrieved from http://www.springer.com/series/7286

Walker, S. L., & Fraser, B. J. (2005). Development and validation of an instrument for assessing distance education learning environments in higher education: The Distance Education Learning Environments Survey (DELES). *Phenomenology and the Cognitive Sciences*, *4*(3), 289–308. https://doi.org/10.1007/s10984-005-1568-3

Widarjono, A. (2010). *Analisis Statistika Multivariat Terapan* (Ppertama). Unit Penerbit dan Percetakan Sekolah Tinggi Ilmu Manajemen YKPN.

Wijsman, L. A., Saab, N., Schuitema, J., van Driel, J. H., & Westenberg, P. M. (2019). Promoting performance and motivation through a combination of intrinsic motivation stimulation and an extrinsic incentive. *Learning Environments Research*, *22*(1), 65–81. https://doi.org/10.1007/s10984-018-9267-z

Yeung, A. S., & Mclnerney, D. M. (2005). Students' school motivation and aspiration over high school years. *Educational Psychology*, *25*(5), 537–554. https://doi.org/10.1080/01443410500046804