

Validation of an Instrument to Measure the Feedback Conceptions Scale

Nor Hasnida Che Md Ghazali, Mahizer Hamzah, Norazilawati
Abdullah & Syaza Hazwani Zaini

Faculty of Human Development, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim,
Perak, Malaysia

Email: hasnida@fpm.upsi.edu.my

To Link this Article: <http://dx.doi.org/10.6007/IJARBS/v9-i7/6091> DOI:10.6007/IJARBS/v9-i7/6091

Published Date: 23 July 2019

Abstract

It is very important to check for validity and reliability of self-report scales or measures. This study aimed to determine the validity and reliability of an instrument which is adapted from Harris and Brown (2008) known as 'Teachers' Conceptions of Feedback (TCof) inventory'. The instrument consists of 37 items from 9 constructs namely irrelevance, improvement, accountability, encouragement and self-type, task, process, self-regulation, peer and self-assessment and timeliness. One hundred and eighty-five (N=185) undergraduate students who are also the student teachers had completed the questionnaire. Respondents used a six-point agreement rating scale coded as strongly disagree=1, mostly disagree=2, slightly agree=3, moderately agree=4, mostly agree=5 and strongly agree=6. The content was validated by three experts in the field of measurement and evaluation. Data analysis was completed using the Rasch measurement model. The findings revealed that most items fit the model as their MNSQ values are between 0.50 and 1.50 except for the three items. Only two items showed a negative point measure correlation indicated that overall, the item discrimination is good. Item reliability and item separation is 0.98 and 7.29 respectively, while person reliability and person separation is 0.89 and 2.81 respectively. In total, four items were deleted altogether. The statistical analysis provides strong evidence to support the validity and reliability of the scale. Therefore, this instrument could be adapted or adopted by other researchers in the Malaysian educational context. Implications of the study is also discussed.

Keywords: Feedback, Validity, Reliability, Self-Assessment, Peer-Assessment

Introduction

Feedback is a very important element in assessment for learning (Hattie, 2009) which could then contribute to an effective teaching. By definition, feedback is an information provided by teachers, peers, parents, books, self or experience regarding one's understanding aspects (Hattie and Timperly, 2007). Information could be in the form of a corrective information, alternative strategy, a clarification of ideas, an encouragement or the correctness of a response. In other word, feedback represent a consequence of one's performance or

understanding. Feedback is to provide knowledge and skills or to develop attitudes amongst students. Feedback is so powerful so much so it is known as among the most critical influences on student learning. The misconception that is going around feedback is always seen as an add-on to teaching. This is not correct. Feedback is supposed to be integrated into the teaching and learning process (Rayment, 2006). Sometimes in reality, assessment is perceived as competitive to the teaching and learning process, whereas assessment is supposed to be the bridge between teaching and learning.

In general, one's conceptions are the beliefs, attitudes and intentions that one's has (Brown, 2008). So, student teachers' conceptions in this study is referring to the student teachers' beliefs, attitudes and intentions towards feedback during assessment process during teaching. It is believed that the conceptions of student teachers contribute a lot to their behavior of them (Ajzen, 2005). A feedback has to be constructive, else it will not work in improving learning. A constructive feedback has to be clear in terms of its goal and standard, information must be in a descriptive phrase, action and solution oriented and strictly confidential so ensure trust is maintained between students and teachers (Black, 2007). But, the most important thing is, a feedback that can make students think is actually the best feedback ever (Leahy et al., 2005). And, in order to make students think, teachers or educators should avoid giving grades, scores or simple comments like 'good job' but teachers are encouraged to give clear and exact comments, and comments given must be related to the rubrics of the syllabus.

There are seven principles of good feedback practice (Nicol and Macfarlane-Dick, 2006). A good feedback serve an opportunity for the students to realize the gap between what is known and what is to be learned, to improve students' self-esteem, to help them to understand their learning goals, to help them in improving dialogue activities with the teachers and also to help teachers to improve their teaching strategies. Teachers are encouraged to implement constructive feedback in their teaching, but many teachers feel that they have not got enough knowledge on that. There are several types of feedback. These categories will have different impact to students' learning (Wiliam, 2011). Starting from the weakest form of feedback whereby students are given mark or grade only. The second is when feedback interventions are in the form of feedback only, which teachers give mark or grade plus feedback on correct answers. This were then followed by a weak formative assessment whereby students are given information on correct answers and some explanation on that. Better than that is when teachers give information on correct answers, some explanations and specific suggestions for improvement. This is called moderate formative assessment. Lastly, which serves as a strong formative assessment or the best feedback intervention is when teachers give information on correct answers, some explanation and specific activities for improvement.

Different kinds of feedback interventions could influence students' learning as shown in Table 1 (Wiliam, 2011). If students are given weak feedback by giving mark or grade only, this would not help students to realize the gap that exists between the current performance and the desired learning goal. However, when they are given information on correct answers together with some explanations and specific activities for improvement, then they will know what to do to improve. Do not forget that our main concern is for the students to make an improvement to their learning.

Table 1

Category of feedback interventions

Feedback intervention	Explanation
Weak feedback	Give mark or grade to students
Feedback only	Give mark or grade and feedback on correct answers
Weak formative assessment	Give information on correct answers and some explanation
Moderate formative assessment	Give information on correct answers, some explanation and some specific suggestions for improvement
Strong formative assessment	Give information on correct answers, some explanation and specific activities for improvement

Feedback is the main component of formative assessment (Heritage, 2007) so, no doubt it is very important in improving performance. But, how effective is feedback? There are 74 meta-analyses conducted looking at the effect of various types of feedback to students' achievement (Hattie, 1999). The analyses found that the best way to provide cues or reinforcement to students are by video-assisted, audio-assisted or computer-assisted instructional feedback. And, the least effective is by using praise, punishment or giving extrinsic rewards. Even, there is a doubt whether reward should be considered as feedback or not. Another study by Kluger and DeNisi (1996) on the effect of feedback showed that giving specific goals together with low task complexity or providing information on correct responses rather than incorrect responses seems to be the most powerful feedback. In addition, praise is the other way round. Why? Because praise actually provide less information related to learning.

As teachers' conception and what they believe in may affect the way they practice constructive feedback, there is a need to assess their belief on feedback conceptions especially the student teachers. However, there is no thorough study in Malaysia regarding teachers' conception of feedback, and this study is adopting an instrument from overseas. Hence, the goal of this present study is to determine the validity and reliability of an instrument which is adapted from Harris and Brown (2008) known as 'Teachers' Conceptions of Feedback (TCoF) inventory'. If this instrument is valid and reliable in the Malaysian context, then anyone can use this as an inventory to determine teachers' conceptions of feedback. A reliable and valid scale of teacher conceptions of feedback could also be useful for researchers, educators or policy makers in determining assessment practices.

To date, there is not many instruments to measure teachers' conceptions of feedback in the Malaysian educational context. Furthermore, there is no application of the Rasch measurement model in measuring the psychometric properties of this instrument which has been validated in the New Zealand educational context. There are some instruments concerning alternative assessment and assessment literacy but it is indirectly related to teachers' conceptions on feedback practises in the classroom such as teachers' alternative assessment instrument by Nurfaradilla et al. (2010) and assessment practises among English teachers in Malaysia (Arsaythambi, 2016). There is an instrument developed on assessment literacy level by Juliana and Siti Eisah (2010) and an instrument which assess the assessment

literacy of Malaysia's home economic teachers by Suriani et al. (2016). So, there is a great need to develop and validate an instrument regarding teachers' conception of feedback so that we can use it in our education system. And, the use of Rasch analysis in investigating the quality of items of this instrument seems to be promising.

Objectives of the Study

The purpose of this study was to investigate empirical evidence regarding the validity and reliability of an established questionnaire named TCoF. The Rasch analysis was conducted in six steps which are Rasch Fit statistics, item difficulty measurement, item polarity, unidimensionality, dispersion and reliability.

Methodology

This study uses survey design. A quantitative approach was used to determine the validity and reliability of the instrument. As this is a pilot study, the respondents involved were 185 student teachers studying in a university in their final semester after coming back from teaching practicals in schools all over Malaysia, and they were selected using simple random sampling. The data was analyzed using the Rasch Measurement Model. The instrument, which was adapted from 'TCoF inventory' (Harris and Brown, 2008) consists of 37 items from 9 constructs (Conception-Irrelevance (Students Ignore) – 4 items, Conception-Improvement (Student Use) – 4 items, Conception-Accountability (Expected) – 3 items, Conception-Encouragement + Self Type (Praise) – 6 items, Task Type (Task) – 3 items, Process Type (Process) – 4 items, Self-Regulation Type (SR) – 5 items, Peer & Self (PASA) – 3 items and Timeliness (Prompt) – 5 items). The scales were measured using a 6-point Likert scale ranging from strongly disagree to strongly agree.

Findings

The findings are described based on the objectives of the study in order to determine the validity and reliability of the instrument on the concepts of feedback among respondents. The analysis was conducted to test item fit, item polarity, unidimensionality, items and respondents' reliability and separation index. When the data from 185 respondents were analysed, it was found that the data was not fit. Analysis were made and data from 83 respondents were deleted. Then, this is what's left, the latest findings referring to the 102 respondents.

Item Fit

The item fit statistics (infit MNSQ and outfit MNSQ statistics) estimate provide an information whether it approaches a certain model or not (Bond and Fox, 2003). They are also used to determine whether the items are contributing meaningfully to the measurement of the construct (Linacre, 2011). The recommended acceptable range for infit and outfit MNSQ statistics is 0.50 to 1.50 (Bond and Fox, 2003), or 0.60 to 1.40 (Linacre, 1994). Items within this range are considered productive (Bond and Fox, 2007). If the individual item does not fill the requirements, then the item will be eliminated. Table 2 reveals that most items show good overall fit of the data to Rasch Model. So, most items are retained with three items are deleted as they are not in the range which are item A2, A4 and I37.

Table 2

Item measure (INFIT, OUTFIT) MNSQ and Point Measure Correlation

Item	Measure	Standard Error	INFIT		OUTFIT		PT MEA CORR
			MNSQ	ZSTD	MNSQ	ZSTD	
A1	1.05	.12	1.28	2.1	1.29	2.2	.27
A2	1.66	.13	1.81	5.0	1.86	5.1	.09
A3	3.53	.17	1.37	2.2	1.34	2.1	-.08
A4	3.18	.16	1.78	4.0	1.70	3.7	-.03
B5	-0.38	.13	0.88	-0.8	0.88	-0.8	.50
B6	-0.40	.13	0.72	-2.1	0.71	-2.1	.58
B7	-1.16	.14	0.90	-0.7	0.90	-0.6	.57
B8	-0.65	.14	0.84	-1.1	0.83	-1.2	.45
C9	-1.04	.14	0.81	-1.3	0.81	-1.3	.49
C10	0.82	.12	1.01	0.2	1.06	0.5	.52
C11	-0.42	.13	1.21	1.4	1.19	1.3	.39
D12	0.45	.12	0.70	-2.5	0.71	-2.4	.60
D13	0.18	.12	0.84	-1.2	0.86	-1.0	.55
D14	0.02	.13	0.94	-0.4	0.98	-0.1	.47
E15	0.16	.13	1.14	1.1	1.15	1.1	.61
E16	-0.16	.13	1.04	0.3	1.07	0.5	.63
E17	-1.32	.14	0.88	-0.8	0.87	-0.9	.55
E18	-0.04	.13	1.55	3.4	1.59	3.6	.41
E19	-0.52	.13	1.05	0.4	1.06	0.5	.52
E20	-1.02	.14	0.74	-1.9	0.74	-1.8	.61
F21	-0.63	.13	0.53	-3.8	0.54	-3.6	.59
F22	-0.76	.14	0.62	-2.9	0.63	-2.8	.57
F23	-1.36	.14	0.81	-1.4	0.81	-1.4	.65
F24	-0.63	.13	0.74	-1.9	0.75	-1.8	.61
G25	-0.47	.13	0.50	-4.0	0.52	-3.9	.64
G26	0.07	.14	1.22	1.6	1.24	1.6	.44
G27	0.19	.13	1.16	1.1	1.20	1.4	.49
G28	0.15	.13	0.81	-1.5	0.82	-1.3	.50
G29	-0.50	.13	0.51	-3.9	0.54	-3.7	.65
H30	-0.33	.12	0.74	-1.9	0.75	-1.8	.60
H31	-0.52	.14	0.57	-3.3	0.58	-3.3	.62
H32	-0.30	.13	0.53	-3.8	0.51	-4.0	.53
I33	0.57	.13	1.14	1.1	1.20	1.5	.31
I34	-0.18	.13	0.88	-0.9	0.89	-0.7	.51
I35	1.27	.13	1.48	3.4	1.50	3.5	.35
I36	-0.69	.13	1.03	0.2	1.06	0.4	.40
I37	0.16	.13	1.95	5.6	1.96	5.5	.26
Mean	0.00	.13	0.99	-0.2	1.00	-0.2	
S.D	1.05	.01	0.36	2.5	0.37	2.4	

Item Polarity

All items show positive item discrimination and a pattern which showed a high validity via a positive correlation point size value. Point Measure Correlation (PMC) is a statistical item showing the correlation results between one points (a response choice) with a continuous variable (scores for all candidates in a test). In Rasch statistics, the mean square value of the residual item which is sensitive to the items which have failed to relate to the test scores and point-biserial items with very large values is considered (Wright and Stone, 1979). In Rasch analysis, we use item correlations as an immediate check that the response-level scoring

makes sense. If the observed correlation is negative, something may have gone wrong. The acceptable critical point measure correlation of an item is 0.2 or more (Pray and Popovich, 1985). In addition, a discrimination index of less than 0.2 is weak and more than 0.4 is good. From Table 1, all items show value more than 0.20 except for the three items. Item A2 shows a value of 0.09 which is less than 0.2. Item A3 and item A4 has a negative value. So, all the two items (item A3 and A4) are deleted. Overall, this findings indicate that the item discrimination is very good.

Unidimensionality

Unidimensionality is important as it shows that the instrument is measuring in one dimension, whether the item is understood or not (Wright and Stone, 1979). For Rasch analysis, for a good unidimensionality, the raw variance explained by measures must be more than 40.0% and the unexplained variance in 1st contrast must be less than 15% (Azrilah, 2010). If refer to Linacre (2006), for raw variance explained by measures, a measurement higher than 40% is considered a strong dimension, higher than 30% is considered a moderately strong dimension and those higher than 20% is considered a moderate dimension. For this study, as shown in Table 3, the raw variance explained by measures is 48.0% (higher than 40.0%) and the unexplained variance in 1st contrast is 6.3% (less than 15%). This indicates all items are clear and not confusing and the measurement have a strong dimension.

Table 3:
Standard residual variance (in Eigenvalue units)

	Empirical			Modeled
Total raw variance in observations	71.10	100.0%		100.0%
Raw variance explained by measures	34.1	48.0%		47.8%
Raw variance explained by persons	7.1	10.1%		10.0%
Raw variance explained by items	27.0	37.9%		37.7%
Raw unexplained variance (total)	37.0	52.0%	100.0%	52.2%
Unexplained variance in 1 st contrast	4.5	6.3%	12.1%	
Unexplained variance in 2 nd contrast	3.8	5.4%	10.3%	

Reliability and Separation

Table 4 shows the summary of statistics of person and item measure. It shows the summary of statistics that measure 102 respondents. The respondents' reliability index is 0.89, which indicates high value and is well accepted. It shows that the respondents are stable and consistent when tested with different items but measuring the same constructs. The respondents' separation index is 2.81 which shows that there are five levels of respondents' ability. Next is to look at the items. The items' reliability index is 0.98. This value shows that the reliability value is very high and acceptable (Pallant, 2007). The items' separation index is 7.29 showing that there are 6 to 7 different levels of items' agreement in this study. If the items' separation level is high, therefore, the measurement instruments are considered better as they have different difficulties. A respondent's reliability is explained on a scale of 0

to 1. The respondents' and items' separation index indicates that it is an acceptable value as it is more than 2. The items and respondents' separation index that is higher than 2 is considered good according to Linacre (2019). Meanwhile, Fisher (2007) stated that the value above 2 is moderately good.

Table 4

Summary of person and item measure

Summary of Person Measurement					
		INFIT		OUTFIT	
	Measurement	MSQ	ZSTD	MSQ	ZSTD
Mean	0.43	1.00	-0.1	1.00	-0.1
SD	0.70	0.38	1.6	0.38	1.6
Separation	2.81				
Person Reliability	0.89				
Summary of Item Measurement					
Mean	0.00	0.99	-0.2	1.00	-0.2
SD	1.05	0.36	2.5	0.37	2.4
Separation	7.29				
Item Reliability	0.98				

Discussion

This study aims to validate an instrument in assessing teacher conceptions of feedback. The validation uses the Rasch measurement model. The analysis revealed that all items fit the model as their MNSQ values are between 0.50 and 1.50 unless for the three items. Two items showed a negative point measure correlation. Other items showed a value more than 0.20 so the item discrimination is good. Item reliability is 0.98 and teacher reliability is 0.89 meaning that they were both showed a very good reliability. The mean for person is 0.43 logits with a standard deviation of 0.70 logits, while the item mean is zero with a standard deviation of 1.05. The overall item quality is good with four items are deleted. The analyses yielded evidence that the instrument can be a useful scale to measure assessment concepts among teachers or student teachers. And, although this instrument originated from the western context, it is suitable to be used in Malaysian context. This instrument can be a starting point for further research. However, a more detailed analysis with a larger sample using Item Response Theory model is needed to explore deeper into the psychometric characteristics of each item.

Conclusion

This study aimed to establish the validity and reliability of an instrument designed to assess teachers' perceptions towards feedback conceptions among teachers. Most items fit the model as their MNSQ values are between 0.50 and 1.50 except for the three items. Only two items showed a negative point measure correlation indicated that overall, the item discrimination is good. Item reliability and item separation is 0.98 and 7.29 respectively, while person reliability and person separation is 0.89 and 2.81 respectively. In total, four items were deleted altogether. The findings reached by Rasch have allowed us to conclude that TCoF inventory is reliable and valid in the Malaysian context and could be used for further research. In conclusion, the use of the Rasch measurement model is good in measuring the validity and reliability of the instrument in the Malaysian educational context especially in secondary level. For further research, the use of larger samples from various school levels (primary and secondary schools) and various countries in Malaysia are recommended.

Contribution of the study

The contributions of this study can be conceptualized into theoretically and practically. An important contribution from the study was it offers a significant starting point for an expansion of future theoretical analyses of assessment practices among teachers in Malaysia. The findings provide an insight and expand towards current theoretical and conceptual regarding teachers' conception of feedback including an understanding of peer-assessment and self-assessment concept. It is also the responsibility of the ministry to conduct an in-service or professional program of teachers so that teachers could improve their level of feedback practices in their assessment implementation during teaching and learning process. Other than teachers, the administrators, policy makers, head masters or curriculum designers might be interested with the findings. The findings also offer a significant starting point for an expansion of future research to use this inventory. Although this inventory has been used in developed countries for few years but still it has to be validated to suit with the Malaysian educational context. The Ministry might also be very concerned with the teachers' conception as it is the most influential factors towards teacher practices in assessment for learning.

Acknowledgement

The authors would like to express their utmost gratitude to the Research Management and Innovation Centre, UPSI for the allocation of University Research Grant (2018-0119-106-01)

References

- Arsaythambi, V. & Rosidah, R. (2016). Assessment Practises among English teachers in Malaysian Secondary Schools, *International Journal of Infonomics (IJI)*, 9(4), 1220-1227.
- Aziz, A. A. (2010). Rasch Model Fundamentals: Scale Construct and Measurement Structure. Kuala Lumpur: Integrated Advanced Planning Sdn.Bhd
- Aziz, A. A. (2010). *Rasch Model Fundamentals: Scale Construct and Measurement Structure*. Kuala Lumpur: Integrated Advanced Planning Sdn Bhd.
- Black, P. (2007) *Formative Assessment: Promises or problems?* Available at: <http://www.mantleoftheexpert.com/studying/articles/Paul%20Black2007.pdf/> (Accessed: 6 July 2013).
- Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah: Lawrence.

- Bond, T. G. & Fox, C. M. (2001) *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, N.J.: Erlbaum.
- Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.
- Fisher, W. P. (2007). Rasch Measurement Transaction. *Transaction of the Rasch Measurement SIG American Educational Research Association*, 21(1), 1095.
- Fisher, W.P. Jr (2007). Rasch Measurement Transaction. Transaction of the Rasch Measurement SIG American Educational Research Association. Vol. 21 No.1, p. 1095
- Fisher, W.P. Jr (2007). Rasch Measurement Transaction. Transaction of the Rasch Measurement SIG American Educational Research Association. Vol. 21 No.1, p. 1095
- Fisher, W.P. Jr (2007). Rasch Measurement Transaction. Transaction of the Rasch Measurement SIG American Educational Research Association. Vol. 21 No.1, p. 1095
- Harris, L. R., & Brown, G. T. (2008). *Teachers' Conceptions of Feedback (TCof) inventory*. Unpublished test. Auckland, NZ: University of Auckland, Measuring Teachers' Assessment Practices (MTAP) Project.
- Hattie, J. & Timperly, H. (2007). The Power of Feedback, *Review of Educational Research*, 77(1), 81-112.
- Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.
- Hattie, J. A. C. (1999). *Influences on student learning*. Inaugural Professorial Address, University of Auckland.
- Hattie, J. A. C. (2003). Teachers make a difference: What is the research evidence? Keynote presentation at the Building Teacher Quality: The ACER Annual Conference, Melbourne, Australia.
- Heritage, M. (2007) Formative Assessment: What Do Teachers Need to Know and Do?, *Phi Delta Kappan*, 89(2), pp. 140-145.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254-284.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Towards the understanding of a double-edge sword. *Current Directions in Psychological Science*, 7, 67-72
- Leahy, S., Lyon, C., Thompson, M. and Wiliam, D. (2005) 'Classroom Assessment: Minute by minute, Day by day', *Educational Leadership*, 63(3), pp.1-7.
- Linacre, J. M. (2011). A user's guide to Winsteps & Ministep Rasch-Model computer programs. Program Manual 3.72.1 <http://www.Winsteps.com/a/winsteps.pdf>
- Nicol, D. J. and Macfarlane-Dick, D. (2006) Formative assessment and self-regulated learning: A model and seven principles of good feedback practice, *Studies in Higher Education*, 31(2), pp. 199-218.
- Pallant, J. (2007) *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows*, 3rd ed. New South Wales: Allen and Unwin.
- Rayment, T. (2006) *101 Essential Lists on Assessment*. Kuala Lumpur: Institut Terjemahan Negara Malaysia Bhd.
- Suriani M, Arasinah, K. & Norhayati, A. (2016). Gauging the assessment literacy of Malaysia's home economic teachers: An empirical study, *Malaysian Journal of Society and Space*, 12(3), 130-138.
- Wiliam, D. (2011) 'What is assessment for learning?' *Studies in educational evaluation*, 37, pp. 3-14.

Wright, B. D. & Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. Chicago, IL: Mesa Press.