Validating a four-tier instrument to investigate students' conception of salt hydrolysis

Isnatul Husniah [a], Habiddin [a, *], Muhammad Sua'idy [a], Nuryono[b]

[a] *Department of chemistry, Mathematics and Natural Sciences Faculty, Universitas Negeri Malang, Indonesia*

[b] *SMP Negeri 1 Batauga, Buton Selatan, South East Sulawesi, Indonesia*

*)Corresponding author: habiddin_wuni@um.ac.id

Abstract

This paper outlines the development and validation of a Four-tier instrument in the topic of salt hydrolysis (FTISH). The instrument was developed by using the procedure of FTDICK instrument as published by Habiddin & Page (2019). This study was a part of other separates studies conducted by other groups of researchers including Devita, Habiddin & Suaidy (2019) (in press). This study involved 127 students in total and data collection were carried out two times involving 64 students in the first data collection and 63 students in the second one. The participants of both groups were public secondary school students from Malang. The FTISH consists of 23 questions. The results show that the validity of the questions ranged between 0,263 and 0,745. The reliability of the instrument was 0,80 and falls in the high category. The future implementation of the instrument to support the teaching and learning of salt hydrolysis are also discussed.

Keywords: four-tier instrument; salt hydrolysis; students' conception, four-tier instrument.

A. Introduction

Students' misunderstanding could be uncovered in many ways [2] including interview, concept map, and multiple-choice. Such those instruments have been applied in many previous studies with some limitation admitted by the authors at the end of their studies. Taking those drawbacks into account, this study developed a multi-tier instrument in order to produce a more reliable and valid instrument to uncover students' conceptions. The use of multi-tier instrument in the area of science and chemistry education discipline was initiated by Treagust from Curtin University, Australia. Following his work, several multi-tier instruments developed and applied in the area in the last two-decades including three-tier instrument [2]–[4]. In considering that the four-tier is the most recent instrument and approved to overcome the drawbacks of the previous multi-tier instrument, this study developed a four-tier instrument in the topic of salt hydrolysis

(FTISH). A four-tier instrument consists of 4 tiers which are tier 1 (multiple choice questions), tier 2 (confidence rating of tier 1), tier 3 (the reason for tier 1) and tier 4 (confidence rating of tier 3) [1].

Acid-base concepts has been a topic putting students into difficulty. Romine, Todd, & Clark (2016) found that students exhibited unscientific understanding in the topic of pH and pH scale. Galuh [6] found that the number of students who got into difficulty in this topic ranged from 65% to 86%. Students experienced a difficulty with very fundamental concepts of acid-base chemistry reaction [7]. Students' alternative conceptions in this field have also been reported in many studies.

B. Material and methods

As stated in the previous section, this study is part of a big research involving some research group. Those groups including Devita, Habiddin & Suaidy (2019) (in press) conducted the same study in the 6 different scholls and participant groups. Later, the result of these studies will be combined to do a single further study. This study involved two group of students which are 63 and 64 students of two different public schools in Malang. Both schools are located in central city of Malang and categorized as urban area schools. The first group was involved in the first data collection using the preliminary instrument. The last group was involved in the second data collection using the four-tier instrument.

The procedure of this study adopted the one developed by Habiddin & Page (2019) including mapping concept, testing & interviewing, defining students' unscientific ideas, developing the prototype four-tier, validating the prototype four-tier and refining the final four-tier.

C. Results and discussion

The empirical validity of FTISH was measured in five criteria including validity, reliability, difficulty level, discriminatory index and distractor effectiveness. The validity index of each question are displayed in Table 1 below.

Tabel 1. Empirical Validity of questions in the FTISH

Q	Answer *Tier* (A)		Reason *Tier* (R)		Both *Tier* (B)	
	r_{xy}	category	r_{xy}	category	r_{xy}	category
1.	0,571	Valid	0,487	Valid	0,591	Valid
2.	0,506	Valid	0,297	Valid	0,471	Valid
3.	0,275	Valid	0,260	Valid	0,267	Valid
4.	0,308	Valid	0,308	Valid	0,376	Valid
5.	0,460	Valid	0,445	Valid	0,560	Valid
6.	0,265	Valid	0,513	Valid	0,332	Valid
7.	0,345	Valid	0,431	Valid	0,363	Valid
8.	0,581	Valid	0,392	Valid	0,430	Valid
9.	0,275	Valid	0,367	Valid	0,363	Valid
10	-0,180	Invalid	0,330	Valid	0,049	Invalid
11	0,355	Valid	0,565	Valid	0,384	Valid
12	0,284	Valid	0,263	Valid	0,349	Valid
13	0,528	Valid	0,255	Valid	0,487	Valid
14	0,067	Invalid	0,088	Invalid	0,063	Invalid
15	0,249	Valid	0,266	Valid	0,272	Valid
16	0,489	Valid	0,298	Valid	0,370	Valid
17	0,261	Valid	0,162	Invalid	0,239	Invalid
18	0,00	Invalid	0,288	Valid	0,283	Valid
19	0,254	Valid	0,066	Invalid	0,263	Valid
20	0,280	Valid	0,484	Valid	0,703	Valid
21	0,642	Valid	0,295	Valid	0,594	Valid
22	0,612	Valid	0,694	Valid	0,745	Valid
23	0,689	Valid	0,664	Valid	0,725	Valid

The table above shows that mostly the questions were valid. When each tier compared, the number of invalid question in each tier is equal. Question 14 (Q14) should be removed as all the tiers were invalid. Q10 is also removed due to the A tier and B tier were invalid. For Q17, Q18 and Q19, a revision was carried out as the invalid one is excited in the R tier only.

Table 2. The reliability of the FTISH

Reliability index					
Answer *Tier* (A)		Reason *Tier* (R)		Both *Tier* (B)	
r_{xy}	Category	r_{xy}	Category	r_{xy}	Category
0,748	Average	0,726	Average	0,800	Tinggi

Table 2 depicts that the reliability of the FTISH is acceptable for all the tiers. The highest index is shown by the B tier.

Table 3. The difficulty level of the FTISH

Category	Difficulty level		
	Answer *Tier* (A)	Reason *Tier* (R)	Both *Tier* (B)
difficult	0	0	0
average	13	6	14
easy	10	17	9

Table 3 shows that none of the questions were considered difficult. The average category ranged equally the A and the B tiers. Surprisingly, the R tier is the one with the ighest number of question considered easy.

Tabel 4. The discriminatory indices of the FTISH

	Answer *Tier* (A)		Reason *Tier* (R)		Both *Tier* (B)	
	Discriminatory index	Category	Discriminatory index	Category	Discriminatory index	Category
1	0,44	Good	0,19	Average	0,47	Good
2	0,44	Good	0,13	Average	0,34	Good
3	0,13	Average	0,28	Average	0,19	Average
4	0,16	Average	0,09	Poor	0,19	Average
5	0,47	Good	0,28	Average	0,53	Good
6	0,25	Average	0,34	Good	0,28	Average
7	0,25	Average	0,25	Average	0,22	Average
8	0,47	Good	0,34	Good	0,44	Good
9	0,16	Average	0,38	Good	0,28	Average
10	-0,16	Inappropriate	0,19	Average	-0,03	Inappropriate
11	0.25	Average	0,38	Good	0,31	Good
12	0,38	Good	0,19	Average	0,25	Average
13	0,25	Average	0,06	Poor	0,19	Average
14	-0,03	Inappropriate	0,00	Inappropriate	0,00	Inappropriate
15	0,16	Average	0,19	Average	0,19	Average
16	0,19	Average	0,16	Average	0,13	Average
17	0,19	Average	0,09	Poor	0,19	Average
18	0,00	Inappropriate	0,38	Good	0,25	Average
19	0,28	Average	0,09	Poor	0,28	Average
20	0,19	Average	0,34	Good	0,69	Good
21	0,47	Good	0,28	Average	0,56	Good
22	0,63	Good	0,63	Good	0,72	Good
23	0,69	Good	0,63	Good	0,69	Good

The discriminatory indices of questions as displayed in Table 4 in each tier were mostly good. Q14 fall in the inappropriate category meaning that the question should be removed.

Table 5. The distractor effectiveness of the FTISH instrument

	Distractor effectiveness									
	1		2		3		4		5	
Opti	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R
A	10,9%	0,0%	15,6%	81,3%	10,9%	12,5%	4,7%	89,1%	45,3%	1,6%
B	12,5%	89,1%	12,5%	14,1%	1,6%	23,4%	0,0%	3,1%	3,1%	81,3%
C	6,3%	6,3%	9,4%	1,6%	21,9%	60,9%	9,4%	1,6%	48,4%	14,1%
D	70,3%	4,7%	62,5%	3,1%	65,6%	3,1%	85,9%	6,3%	3,1%	3,1%
	6		7		8		9		10	
Opt	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R
A	6,3%	4,7%	4,7%	9,4%	17,2%	1,6%	3,1%	14,1%	17,2%	9,4%
B	7,8%	73,4%	9,4%	10,9%	64,1%	12,5%	3,1%	0,0%	65,6%	3,1%
C	81,3%	15,6%	68,8%	3,1%	0,0%	57,8%	3,1%	3,1%	3,1%	0,0%
D	4,7%	6,3%	17,2%	76,6%	18,8%	28,1%	89,1%	82,8%	14,1%	88%
	11		12		13		14		15	
Opt	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R
A	28,1%	81,3%	6,3%	37,5%	84,4%	0,0%	90,6%	1,6%	51,6%	12,5%
B	6,3%	0,0%	45,3%	7,8%	1,6%	0,0%	0,0%	71,9%	14,1%	15,6%
C	64,1%	15,6%	6,3%	46,9%	0,0%	3,1%	7,8%	7,8%	26,6%	50,0%
D	1,6%	3,1%	42,2%	7,8%	14,1%	96,9%	1,6%	18,8%	7,8%	21,9%
	16		17		18		19		20	
Opt	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R	tier A	tier R
A	1,6%	3,1%	71,9%	7,8%	0,0%	9,4%	40,6%	78,1%	7,8%	0,0%
B	85,9%	79,7%	7,8%	76,6%	1,6%	0,0%	20,3%	6,3%	85,9%	14,1%
C	6,3%	17,2%	7,8%	4,7%	96,9%	82,8%	25,0%	15,6%	6,3%	70,3%
D	6,3%	0,0%	12,5%	10,9%	1,6%	7,8%	14,1%	0,0%	0,0%	15,6%
	21		22		23					
Opt	tier A	tier R	tier A	tier R	tier A	tier R				
A	1,6%	3,1%	4,7%	53,1%	59,4%	4,7%				
B	64,1%	7,8%	42,2%	37,5%	35,9%	32,8%				
C	12,5%	71,9%	3,1%	3,1%	1,6%	6,3%				
D	21,9%	17,2%	50,0%	6,3%	3,1%	56,3%				

Table 5 shows that all the option were selected by some students. This implies that all the distractors were effective. However, some options need to be revised in considering the low number of students that selected it and with regards of other parameters.

D. Conclusions

This study found that the FTISH is a valid and reliable instrument for identifying students' conceptions of acid-base concepts.

References

[1]   H. Habiddin and E. M. Page, "Development and Validation of a Four-Tier Diagnostic Instrument for Chemical Kinetics (FTDICK)," *Indones. J. Chem.*, vol. 19, no. 3, pp. 720–736, May 2019.

[2]   A. C. Dindar and O. Geban, "Development of a three-tier test to assess high school students' understanding of acids and bases," in *3rd World Conference on Educational Sciences*, vol. 15, 2011.

[3]   H. Pesman and A. Eryilmaz, "Development of a Three-Tier Test to Assess Misconceptions About Simple Electric Circuits," *J. Educ. Res.*, vol. 103, no. 3, pp. 208–

222, 2010.

[4]   D. Milenkovic, D. M. Dusica, N. H. Tamara, D. S. Mirjana, and H. Sasa, "Development of a Three-Tier Test as a Valid Diagnostic Tool for Identification of Misconceptions Related to Carbohydrates," *J. Chem. Educ.*, vol. 93, no. 9, p. 1514, 2016.

[5]   W. L. Romine, A. N. Todd, and T. B. Clark, "How Do Undergraduate Students Conceptualize Acid–Base Chemistry? Measurement of a Concept Progression," *Sci. Educ.*, vol. 100, no. 6, pp. 1150–1183, Nov. 2016.

[6]   G. Febriani, S. Marfu'ah, and R. Joharmawan, "Identifikasi Konsep Sukar, Kesalahan Konsep, dan Faktor-Faktor Penyebab Kesulitan Belajar Hidrolisis Garam Siswa Salah Satu SMA Blitar," *J-PEK (Jurnal Pembelajaran Kim.*, vol. 3, no. 2, pp. 35–43, 2018.

[7]   J. M. Nyachwaya, "General chemistry students' conceptual understanding and language fluency: acid–base neutralization and conductometry," *Chem. Educ. Res. Pract.*, vol. 17, no. 3, pp. 509–522, 2016.