STEM Education International,
Vol. 1, No. 2, December 2025, Page 74-79
p-ISSN: xxxx-xxx, e-ISSN: xxx-xxx

STEM-EI
STEM Education International

Development of a Four-Tier Diagnostic Instrument to Identify High School Students' Understanding of Salt Hydrolysis

Anne Nailul Aziz, Darsono Sigit

Department of Chemistry, Universitas Negeri Malang, Indonesia 65145

**Corresponding author: annelula9@gmail.com*

Abstract: This research aims to (1) produce a four-tier instrument in salt hydrolysis, (2) determine the validity and reliability of the instrument to facilitate students' understanding of salt hydrolysis. The development of the instrument followed the platform developed by Habiddin & Page (2019), which was adapted from the Treagust (1988), which consisted of six steps. The results of instrument validation were 81.74% with a very feasible category, yielding 23 valid questions, with a reliability of 0.7985.

Keywords: understanding, four-tier diagnostic instrument, salt hydrolysis

INTRODUCTION

Chemistry is developed through experiments to answer questions about what, why, and how natural phenomena, particularly those related to the composition, structure, properties, transformations, dynamics, and energetics of matter. Salt hydrolysis is a chemistry topic studied by 11th-grade science students in high schools, particularly in Indonesia (Amala & Habiddin, 2022; Habiddin et al., 2022). Students must not only acquire knowledge but also engage in critical and creative thinking (Nafiah et al., 2025). Therefore, efforts to uncover students' deep understanding help inform the design of proper chemistry teaching. The process of identifying misconceptions can be done using diagnostic tests. Diagnostic tests are used to determine the cause of students' learning failures. A diagnostic test is a test used to identify weaknesses (misconceptions) in specific topics and to provide feedback on students' responses to improve their performance. A four-tier format has been used in many chemistry studies for this purpose, including chemical equilibrium (Tyson et al., 1999), chemical bonding (Amalia & Habiddin, 2024; Peterson et al., 1989; Tan & Treagust, 1999), qualitative

analysis (Tan et al., 2002), acid-base properties of salt solutions (Habiddin et al., 2021), thermodynamics (Sreenivasulu & Subramaniam, 2013), metal transition (Sreenivasulu & Subramaniam, 2014), chemical kinetics (Habiddin & Page, 2023; Yan & Subramaniam, 2018) and other topics. The four-tier diagnostic test is an extension of the three-tier multiple-choice diagnostic test, adding a confidence level for each answer and reason (Caleon & Subramaniam, 2010). Adding a confidence rating to each answer and reason can measure differences in students' knowledge levels and help detect the extent of their misconceptions. The four-tier diagnostic test was developed to determine how well students have mastered concepts by measuring their confidence in answering questions.

The first tier of the four-tier diagnostic test consists of multiple-choice questions with three distractors and one correct answer that students must select. The second tier is the students' confidence level in determining their answers. The third tier is the reason students answered the question, consisting of three pre-defined reason options and one open-ended reason. The fourth tier is the students' confidence level in selecting the reason (Habiddin & Nofinadya, 2021; Habiddin & Page, 2019).

METHOD

The instrument development in this study adapted the 6-stage procedures, including (1) concept mapping, (2) testing and interviewing, (3) defining students' unscientific ideas, (4) developing the four-tier prototype, (5) validating the four-tier prototype, and (6) refining the final four-tier instrument (Habiddin & Nofinadya, 2021; Habiddin & Page, 2019). The instrument was evaluated by 2 validators: one lecturer from the Chemistry Department and one chemistry teacher from a public secondary school in Tulungagung, East Java, Indonesia. The initial stage employed open-ended multiple-choice questions and involved 103 students from a public secondary school in Tulungagung. From this mapping, a set of 30 questions was developed and tested with another group of 69 students who had studied salt hydrolysis. The empirical data obtained from students' answers were analysed to determine the validity, reliability, difficulty level, item discrimination, and distractor effectiveness of the four-tier instrument of salt hydrolysis.

RESULTS AND DISCUSSION

A total of 23 test items were identified as valid and reliable. Students choose one answer and one reason that they believe is correct, and they also select their level of confidence in answering the question and providing the reason. The validation was conducted to test the feasibility of the developed four-tier diagnostic instrument and to assess the suitability of the questions, question indicators, and key concepts in the salt hydrolysis material. The content validation results showed that the developed instrument had an average percentage of 81.74%, which falls into the very feasible category (Arikunto, 2021). In empirical validation, a validity level analysis is performed. Based on calculations, the 23 developed questions were declared valid, with r-calculated > r-table at a significance level of 0.05. The results of the item validity level analysis are presented in Table 1.

Table 1. Validity of Items

Soal		1	2	3	4	5	6	7	8	9	10	11	12
A Tier	r_{xy}	0.32	0.62	0.51	0.76	0.75	0.50	0.78	0.36	0.79	0.48	0.87	0.24
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
R Tier	r_{xy}	0.46	0.61	0.23	0.27	0.46	0.49	0.60	0.33	0.64	0.24	0.27	0.28
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
B Tier	r_{xy}	0.39	0.61	0.37	0.51	0.60	0.49	0.69	0.34	0.72	0.36	0.58	0.26
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid

Soal		13	14	15	16	17	18	19	20	21	22	23
A Tier	r_{xy}	0.49	0.38	0.55	0.30	0.48	0.40	0.48	0.43	0.53	0.46	0.28
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
R Tier	r_{xy}	0.37	0.31	0.29	0.47	0.34	0.40	0.59	0.25	0.26	0.52	0.24
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
B Tier	r_{xy}	0.43	0.35	0.42	0.38	0.41	0.40	0.54	0.34	0.39	0.49	0.26
	Category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid

The reliability of the 23 items tested was 0.883 for the A tier (answers tier), 0.714 for the R tier (reason tier), and 0.7985 for the B tier. In the item discrimination power analysis, in the answer tier (A), there is 1 item categorised as poor, 13 items categorised as fair, 5 items categorised as good, and 4 items categorised as very good. In the reason tier (R), there are 7 items categorised as poor, 9 as fair, and 7 as good. In the both tier (B)/in both tiers, there is 1 item categorised as poor, 15 items categorised as fair, and 7 items categorised as good. The item difficulty level analysis showed that 9 items were categorised as easy, 14 as moderate, and 1 as difficult in the answer tier (A). The test results for difficulty level are presented in Figure 1.

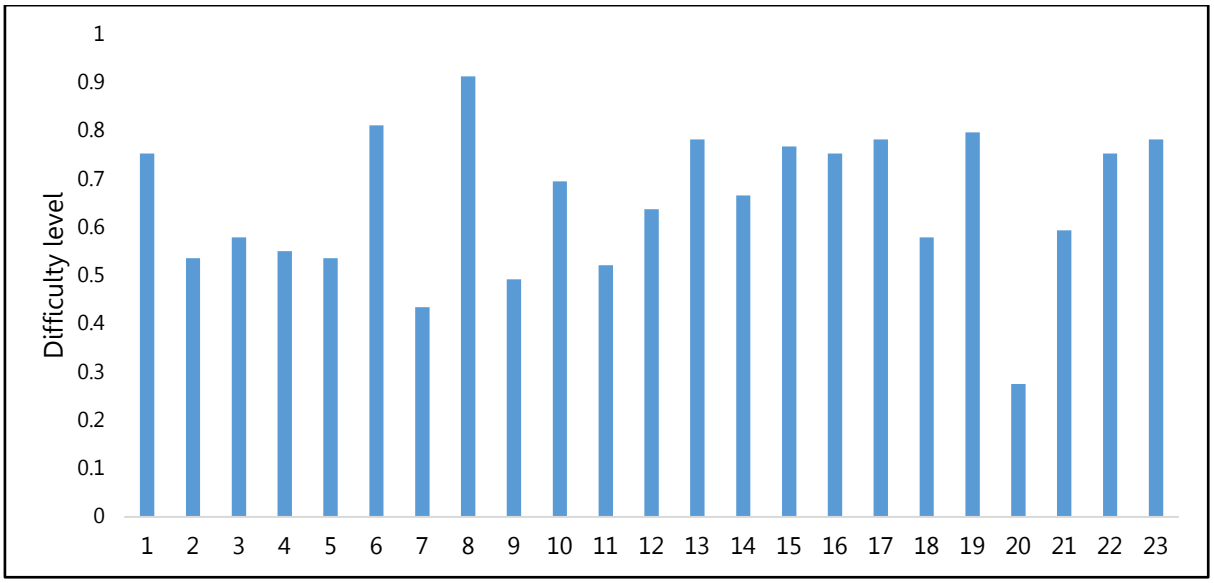

Figure 1. Difficulty indices of Answer Tier (A)

At the reasoning tier (R), 8 questions were considered easy and 15 moderate. The difficulty level test results are presented in Figure 2.

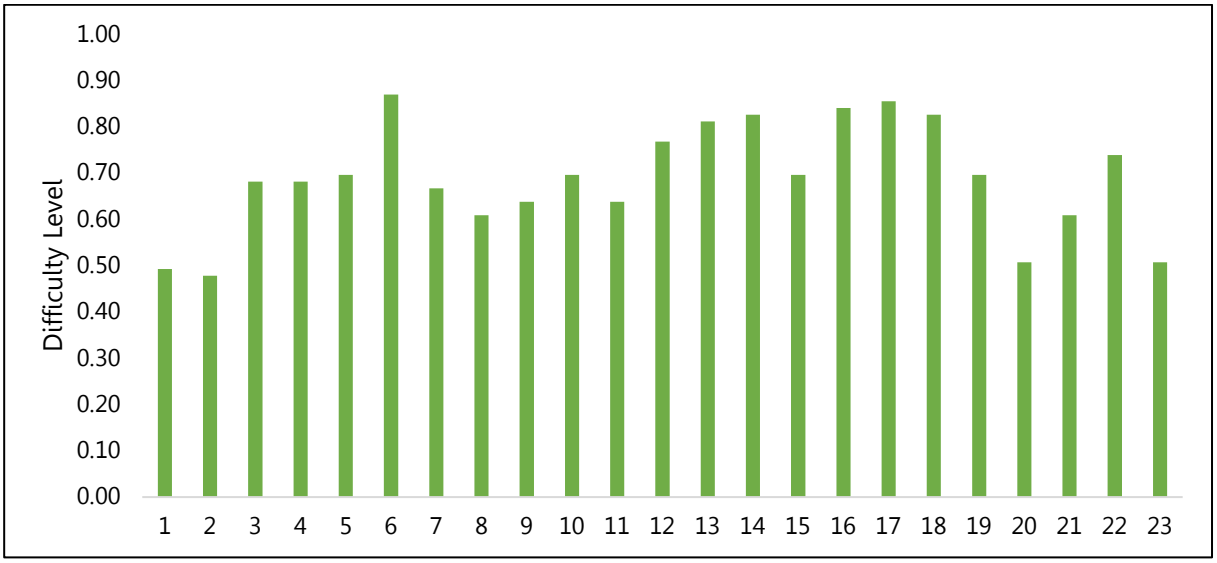

Figure 2. Difficulty indices of Reason Tier (R)

At both tiers (B), there are 9 easy questions, 13 medium questions, and 1 difficult question. The difficulty level test results are presented in Figure 3. The quality of the distractors (their effectiveness) is measured for each multiple-choice question. The criteria for determining whether a distractor is functioning well are met if it is selected by at least 5% of test-takers (Arikunto, 2021). The results of the analysis for each indicator are presented in a table showing the percentage level of each indicator, which represents the analysis of misconceptions occurring for each indicator. The results of the distractor effectiveness calculation are shown in Table 2.

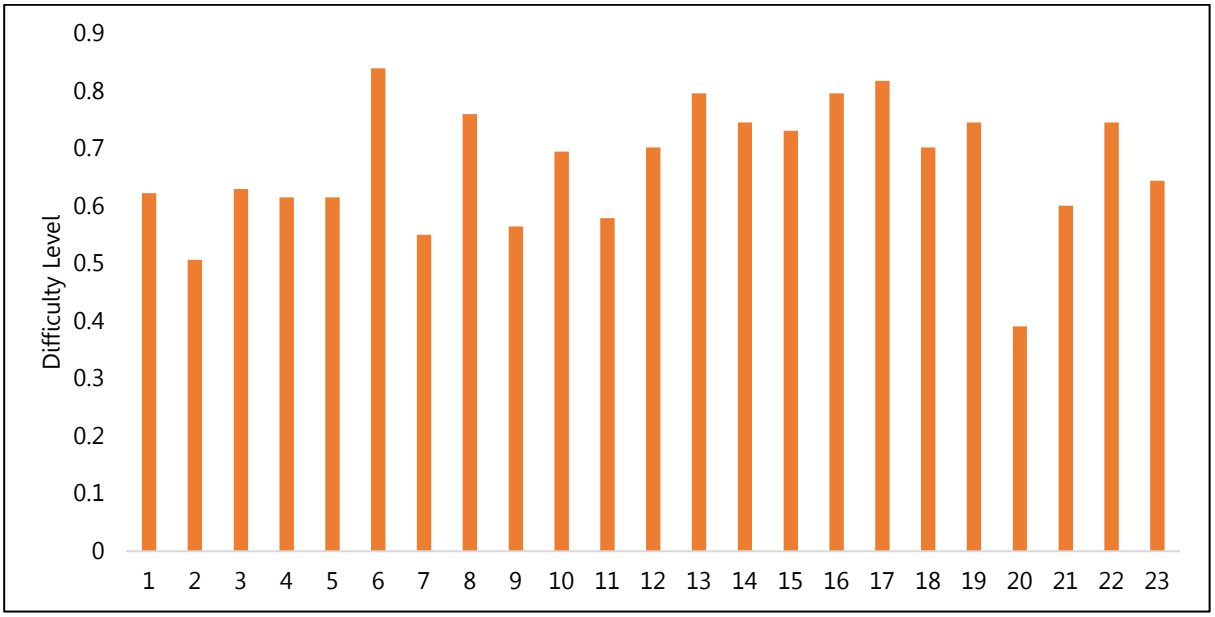

Figure 3. Difficulty indices of Both Tier (B)

Table 2. Distractor Effectiveness of Items

Option	1		2		3		4		5		6		7		8	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	8.70	20.29	53.62	1.45	57.97	8.70	31.88	68.12	31.88	69.57	15.94	8.70	27.54	17.39	0.00	5.80
B	75.36	30.43	24.64	39.13	27.54	68.12	8.70	7.25	14.49	0.00	2.90	2.90	43.48	8.70	2.90	60.87
C	11.59	49.28	2.90	11.59	2.90	15.94	4.35	4.35	0.00	13.04	81.16	86.96	2.90	66.67	5.80	21.74
D	4.35	0.00	18.84	47.83	11.59	7.25	55.07	18.84	53.62	17.39	0.00	1.45	26.09	7.25	91.30	11.59

Option	9		10		11		12		13		14		15		16	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	20.29	13.04	1.45	69.57	52.17	63.77	31.88	0.00	78.26	81.16	66.67	5.80	1.45	69.57	7.25	84.06
B	49.28	63.77	27.54	13.04	15.94	4.35	23.19	13.04	7.25	14.49	11.59	4.35	76.81	13.04	13.04	4.35
C	1.45	5.80	69.57	10.14	7.25	18.84	18.84	10.14	10.14	2.90	2.90	7.25	21.74	4.35	4.35	8.70
D	28.99	17.39	1.45	7.25	24.64	13.04	26.09	76.81	4.35	1.45	18.84	82.61	0.00	13.04	75.36	2.90

Option	17		18		19		20		21		22		23	
	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier	A tier	R tier
A	10.14	1.45	8.70	82.61	5.80	69.57	28.99	10.14	59.42	5.80	75.36	73.91	4.35	2.90
B	78.26	1.45	57.97	4.35	79.71	5.80	27.54	18.84	24.64	13.04	4.35	8.70	78.26	13.04
C	4.35	85.51	1.45	2.90	11.59	4.35	30.43	50.72	10.14	60.87	7.25	5.80	7.25	33.33
D	7.25	11.59	31.88	10.14	2.90	20.29	13.04	20.29	5.80	20.29	13.04	11.59	10.14	50.72

The students' response when completing this instrument was that they had never done diagnostic four-tier model questions before. Therefore, some students still felt confused at the beginning of the test, complaining that the questions were multi-page, which made them less enthusiastic. Before the pilot test is conducted, the researcher must also explain in detail the steps for answering the questions. Additionally, students are unfamiliar with microscopic image questions, leading them to provide less severe answers and prompting them to guess when responding. Some students, when answering the level of confidence in choosing answers and reasons, answered carelessly, either guessing everything or answering with complete certainty. This indicates that some students are not taking the given questions seriously.

CONCLUSIONS

The study developed 23 questions on salt hydrolysis in a four-tier format, with a reliability of 0.79, which falls within the acceptable category. The set of 23 questions was derived from the 30 initial items after applying the validation procedures. All the items were also found to be valid and suitable to identify secondary school students' understanding of salt hydrolysis. The confidence level attached to the reason tier for the instrument uses a 5-point scale (1 = guessing, 2 = unsure, 3 = moderate, 4 = confident, 5 = very confident) as proposed in the previous study (Habiddin & Nofinadya, 2021).

REFERENCES

Amala, F., & Habiddin, H. (2022). Pemahaman konsep dalam topik sifat asam basa larutan garam: studi pada siswa SMA di Blitar. *Jurnal Zarah*, *10*(2), 91–100. https://doi.org/10.31629/ZARAH.V10I2.4321

Amalia, E., & Habiddin, H. (2024). Four-tier chemical bonding instrument: Its development and validation. *Edu Sains: Jurnal Pendidikan Sains Dan Matematika*, *12*(1), 21–35. https://doi.org/ttps://doi.org/10.23971/eds.v12i1.5048

Arikunto, S. (2021). *Dasar-dasar evaluasi pendidikan edisi 3*. Bumi Aksara.

Caleon, I., & 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.

Habiddin, H., Akbar, D. F. K., Husniah, I., & Luna, P. (2022). Uncovering Students'

Understanding: Evidence For The Teaching of Acid-Base Properties of Salt Solution. *Educación Química*, *33*(1), 64–76. https://doi.org/10.22201/FQ.18708404E.2022.1.79488

Habiddin, H., Atikah, A., Husniah, I., Haetami, A., & Maysara, M. (2021). Building scientific explanation: A study of acid-base properties of salt solution. *AIP Conference Proceedings*, *2330*(1), 20047. https://doi.org/10.1063/5.0043215

Habiddin, H., & Nofinadya, S. A. (2021). *The Multi-Tier Instrument in the Area of Chemistry and Science* (M. Bouezzeddine (ed.); p. Ch. 1). IntechOpen. https://doi.org/10.5772/intechopen.100098

Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, *19*(3), 720–736. https://doi.org/10.22146/ijc.39218

Habiddin, H., & Page, E. M. (2023). Uncovering Students' Genuine Misconceptions: Evidence to Inform the Teaching of Chemical Kinetics. *Acta Chimica Slovenica*, *70*(2), 184–195. https://doi.org/https://doi.org/10.17344/acsi.2022.7880

Nafiah, N. I., Damayanti, A., Winarno, A. N. P., Akmalia, D. R., Nurdaningrum, F., Azizah, F. A., Juditha, I., Widya, I. S., Afnia, M. N., & Martiningrum, R. (2025). Learning Approach for Enhancing Students' Creativity. *STEM Education International*, *1*(1), 29–35. https://doi.org/10.71289/wrycf010

Peterson, R. F., Treagust, D. F., & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate grade-11 and -12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, *26*(4), 301–314.

Sreenivasulu, B., & Subramaniam, R. (2013). University Students' Understanding of Chemical Thermodynamics. *International Journal of Science Education*, *35*(4), 601–635.

Sreenivasulu, B., & Subramaniam, R. (2014). Exploring Undergraduates' Understanding of Transition Metals Chemistry with the use of Cognitive and Confidence Measures. *Research in Science Education*, *44*(6), 801–828.

Tan, K.-C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching*, *39*(4), 283–301.

Tan, K.-C. D., & Treagust, D. F. (1999). Evaluating Students' Understanding of Chemical Bonding. *School Science Review*, *81*(294), 75–83.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, *10*(2), 159–169.

Tyson, L., Treagust, D. F., & Bucat, R. B. (1999). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education*, *76*(4), 554–558.

Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chemistry Education Research and Practice*, *19*(1), 213–226.