Development and Validation of a Four-Tier Diagnostic Instrument for Chemical Kinetics (FTDICK)

Habiddin[1,*] and Elizabeth Mary Page[2]

[1]*Department of Chemistry, Faculty of Mathematics and Science, Universitas Negeri Malang, Jl. Semarang No. 5, Malang 65145, East Java, Indonesia*

[2]*Department of Chemistry, University of Reading, Whiteknights Park, Reading RG6 6AD, United Kingdom*

** **Corresponding author:***

tel: +62-341-567382
email: habiddin_wuni@um.ac.id

Received: September 28, 2018
Accepted: November 17, 2018

DOI: 10.22146/ijc.39218

Abstract: *The present study outlines the development and validation of a four-tier diagnostic instrument to explore first-year undergraduate students' understanding of chemical kinetics (FTDICK). The four-tier instrument is a recent format and has been applied only sparsely in a limited number of subject areas, not including chemical kinetics. This study confirms the importance of a four-tier approach in fully investigating students' poor knowledge and understanding. The FTDICK described here involves 20 questions with a confidence level linked to both the question tier and the reason tier. The development of the instrument followed the procedure used for the two-tier instrument developed by Treagust and involves (1) testing & interviewing, (2) identifying & collecting students' unscientific ideas, (3) developing the prototype FTDICK, (4) validating the prototype FTDICK, and (5) developing the final FTDICK. The initial steps revealed a number of areas of misconceptions to be explored in the final instrument. The instrument has been developed, and the prototype tested using international cohorts of students from the University of Reading, UK and two Indonesian Universities. Comprehensive item analysis on the results showed the instrument to be valid and reliable and suitable for identifying students' understanding of chemical kinetics. This study confirms the importance of a four-tier approach for investigating students' prior knowledge and understanding.*

Keywords: *validation; four-tier instrument; Four-Tier Diagnostic Instrument for Chemical Kinetics (FTDICK)*

INTRODUCTION

Prior knowledge, or pre-concepts, of students when they embark on new learning have a significant impact upon their success when undertaking specific studies. The influence of chemistry undergraduate students' prior knowledge on the quality of learning in general chemistry courses has been uncovered by [1-2] and others. However, typically chemistry classes are conducted without first identifying undergraduates' preconceptions. Some lecturers assume their students have mastered some basic chemistry concepts because they have completed the subject in secondary schools. In fact, students continue to harbor misunderstandings and alternative concepts as reported in earlier studies [3]. This mismatch of educator assumptions regarding students' prior knowledge can cause further confusion in the student learner [5]. In addition, lecturers have a limited time frame in which to identify students' pre-concepts comprehensively.

Another study [6] emphasized the importance of research into university chemistry students' prior knowledge for the above reasons and concluded that such alternative concepts would contribute to ineffective learning in future studies. The research underlines the importance of investigating students' understanding before and after learning new concepts. Therefore knowledge of students' alternative concepts can help educators design their teaching more precisely [7].

The term 'unscientific idea' refers to an idea held by students which is incorrect/unscientific and not accepted by the scientific community. In chemistry and science education literature, the terms 'misconception and 'alternative conception' [8] have been used to mean an 'unscientific idea,' and these terms are used throughout this paper. The terms 'misconception' and 'alternative conception' are interchangeable.

The Historical Development of Multiple-Tier Instruments

Students' understanding of chemical concepts, especially misunderstandings and alternative concepts, cannot be fully identified by a simple instrument such as a multiple-choice test. Such one-dimensional tools cannot distinguish between a lack of knowledge, an alternative concept or indeed guesswork. A diagnostic instrument is preferred to identify students' understanding and scientific ideas more comprehensively. Other approaches for exposing student understanding have included interviews [9], concept mapping [10], open-ended questions as well as multiple-choice test [11-12]. Each of these methods has limitations. Using concept mapping requires students to be able to master vocabulary in order to express their ideas in a logical manner [13], while interviews are time-consuming [14]. Open-ended questions can be informative in exposing students' thought processes and procedural applications, but workable sample sizes limit the transferability of the results to whole cohorts. In multiple-choice tests, the validity of the items and test reliability values can be affected by students' test-wiseness skills [15] and cannot always uncover the full reasons for students' answers [16]. Guesswork is a major factor which can impact on the results in a single-tier multiple choice test [17].

Multiple-tier diagnostic tests have become more widely used instruments in science education research recently. Treagust [18] initiated the use of two-tier diagnostic instruments which are the most popular form of multi-tier instruments. The first tier consists of multiple-choice questions with one correct answer and several controlled distractors. The second tier is a reason tier which can be in two different formats. In one format the second tier consists of choices for a reason for the chosen answer in multiple-choice question (MCQ) form. This tier includes one correct scientific reason and several incorrect or unscientific reasons [14,19]. All options in both tiers are derived from students' actual misconceptions and misunderstandings as determined by preliminary tests, interviews and the literature. The tier displaying the reasons should be constructed from students' ideas and concepts determined through an interview or other means [18]. In the second type of two-tier instrument, the reason tier is open-ended. This format has seen less useful as it is more difficult to code the responses; however, it can be extremely revealing as a diagnostic tool on which to base interview questions.

This type of instrument has been used to identify students' understandings in several topics, for example, covalent bonding [11,20], chemical equilibrium [21], and qualitative analysis [19]. However, there are deficiencies in this type of two-tier instrument that have been articulated in the last decade. In some cases, students select the reason for their answer randomly, and such a response does not imply that the student harbors that particular misconception, but rather they are not certain why they have selected their answer in the first tier. A two-tier instrument of this type cannot distinguish between a guessed reason and a real misconception [22]. In addition, a two-tier instrument could only detect a small proportion of students' alternative concepts [23]. Recently, three-tier and four-tier instruments have been used in science education research to overcome the drawbacks of the two-tier instrument. In both types of instrument, the first two tiers are exactly the same as the tiers in the two-tier instrument. However, the three-tier instrument asks students for their mean confidence level in both tiers while the four-tier instrument asks for the confidence level in both tiers separately [4]. Confidence levels are rated on a scale from 1(not at all confident) to 5 or 6 (absolutely confident).

As the confidence level in the three-tier instrument relates to mean confidence in both answer and reason, such an instrument cannot determine whether students are confident about both their answer and their reason or whether they are sure about their answer but unsure about their reason and vice versa. This creates difficulties when grading and categorizing the responses [24]. To overcome this disadvantage of the three-tier instrument, the four-tier instrument has been developed. A confidence level rating is added to students' answers both the multiple-choice questions (the first tier) and the reason tier as proposed by Arslan et al. [24] and Loh et al. [25]. When a level of confidence is attached to both the answer and the reason a greater certainty about understanding and guesswork can be developed. A three-tier instrument has the limitation that the answer tier and reason tier may hold different conceptual problems for a student whereas a four-tier instrument allows them to separately express their confidence in the answer from their confidence in their reason. A student with a good understanding of how to do the problem and why it is correct should display a high confidence level in both tiers. A student with low confidence in their answer but high confidence in their reason may remember the theory correctly but not be able to apply it. Clearly other combinations can be explored which lead to a deeper picture of student understanding of a particular concept.

In our study, we have chosen to develop a four-tier diagnostic instrument in order to explore first-year undergraduate students' understanding in chemical kinetics. Since the four-tier instrument was introduced in 2010, it has not been widely used to date [26], and studies on first-year undergraduate students in this subject area are timely [27].

This instrument will provide both theoretical and practical advantages. Theoretically, this instrument will enrich chemistry education literature and will be transferable to related research. Practically, this instrument will help university lecturers in their general and physical chemistry classes as it can be used to identify students' preconceptions and so assist in planning to teach. So far, studies using a four-tier instrument have only been published in physics education [28], in chemical thermodynamics [26] and transition-metal chemistry [29] and none of these studies have focused on chemical kinetics.

A Four-Tier Instrument in Chemical Kinetics (FTDICK)

Chemical kinetics is a topic that has been of concern to chemical education researchers at the undergraduate level for some time. It has been reported that university students get into difficulty when understanding certain concepts in chemical kinetics even if they have covered the topic in secondary school. For example, students' assumptions that the order of reaction is directly related to the coefficient of a reactant in the chemical equation shows how students are not aware that the order of reaction is determined experimentally [30]. A recent review concluded there is a strong need for a tertiary-level investigation into the teaching and learning of chemical kinetics [27]. The most recent study in this area involved grade 12 students (secondary level) and deployed a three-tier instrument [31]. Chemical kinetics is one of the most important topics in the chemistry curriculum, whether at secondary or university level and is also of relevance to other disciplines such as pharmacy, biochemistry, food science, etc. In many degree programmes, chemical kinetics is taught throughout the course being covered in general or fundamental chemistry in the first year, in physical chemistry in the second year and in chemical kinetics in the third year. The course arrangement can be different from one department to another and can vary internationally. In the UK physical chemistry is generally taught initially in year 1 whereas many US and Asian courses include a preliminary year of general chemistry. Concepts in chemical kinetics play a significant role in explaining the relationships between energy and chemical change, chemical reaction types, and the process of chemical changes [32].

Part 1 of our study describes the development and validation of an FTDICK by adapting the procedure for a two-tier instrument as defined by Treagust [18]. This procedure involves initial testing and interviewing, paper-and-pencil test and four-tier test development validation. The purpose of the testing step is to identify and collect students' unscientific ideas and alternative concepts in the subject area based on their answers to initial multiple-choice (MC) and open-ended answer tests. Based on the responses several students were selected for follow-up interview sessions in order to confirm and explore the more unusual findings in their responses to the tests. This step is important as it provides experiential material on which to design distractors and unscientific reasons in the FTDICK instrument. In addition, unscientific ideas from the literature and from the authors' experiences in teaching the subject are also used in the instrument design. Once a FTDICK has been produced the next step is to test the content and validity using various statistical procedures and to refine the final instrument before carrying out a pilot study on a student sample.

The main purposes of this study are to develop and validate a four-tier diagnostic instrument in chemical kinetics (FTDICK) in order to contribute to the teaching and understanding of the subject the university level by providing an effective and efficient instrument that can be implemented by chemistry educators. Specifically, the research question addressed in this study is "is the FTDICK a valid and reliable instrument to investigate students' understanding of chemical kinetics?". Once proven to be valid and reliable the instrument will be used with a large international sample of undergraduate students to explore their understanding of the area of chemical kinetics.

■ EXPERIMENTAL SECTION

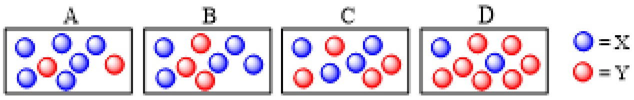
The development of an FTDICK followed the procedure developed by Treagust from Curtin University with some changes. The development includes several steps: (1) testing and interviewing students to obtain a general overview of concepts that are not generally understood; (2) collecting and categorizing students' unscientific ideas and misconceptions based on initial questionnaire results and interviews in (1); (3) developing a prototype FTDICK based upon the findings from (1) and (2); (4) validating the prototype FTDICK by piloting with sample students; and (5) refining the final FTDICK based on findings from (3).

The main purpose of steps 1 and 2 was to identify the chemical kinetics concepts that will be explored in the study. This involves ensuring students have studied the concept to be investigated and that the concept is relevant to the 1st year undergraduate curriculum. Based on inspection of various chemistry curricula, including A-Level and GCSE chemistry (in the UK), the secondary school chemistry syllabus in Indonesia and basic chemistry textbooks, several concepts were defined. These include half-life, the dependence of rate on concentration and temperature, the order of reaction, activation energy, collision theory and the rate law or rate equation. The aims of this initial concept identification are to define the scope of content that will be explored and the level of questions that should be provided.

In order to construct the questions, various sources were exploited including textbooks and exam papers. All questions were original and developed by the authors. The prototype questions were then scrutinized by colleagues who are physical chemists responsible for teaching chemical kinetics to final year undergraduate level and a small number of the questions adjusted as a result. In this way 20 MCQ's were written, and an example of one such question is given in Fig. 1. The item consists of a question with one correct answer (B) and 3 distractors. Following the question, students were asked to give their level of confidence in their chosen answer using a 5 part Likert scale from very unconfident to very confident. Students were then asked to select the appropriate reason for their answer given in the first tier and then assign a level of confidence to the reason, again using a five-part Likert scale. Reason options A, B, and D, are all incorrect and were adopted from students' unscientific ideas as uncovered by the preliminary study.

The example in Fig. 1 uses a pictorial device to represent the concentrations of reacting molecules. Both pictorial questions and more typical numerical representations were used in the MCQ. In some cases, the same question type was displayed in both representations to determine whether either question type led to more informed responses.

For a hypothetical reaction: X + Y → Products, the rate of reaction is second order with respect to X but first order with respect to Y. Four experiments are carried out with different starting concentrations represented pictorially below in the boxes A, B, C and D. Which of the starting conditions (A, B, C or D) will result in the highest rate of reaction?



State the confidence rating of your answer

1. Very unconfident 2. Not very confident 3. Average 4. Quite confident 5. Very confident

Which one of the following options is the reason for your answer to the question?

A. It has the highest concentration of the reactant which is 2nd order
B. The concentrations of both reactants are the same, therefore the ratio of collision is more favorable
C. The amount of X and Y and the average of each keep determine the rate
D. The concentration of Y is much higher than the concentration of X and this leads to the reaction being completed faster

State the confidence rating of your answer

1. Very unconfident 2. Not very confident 3. Average 4. Quite confident 5. Very confident

Fig 1. Example of an item in the prototype FTDICK instrument

Testing and Interviewing

This step aims to verify the level and breadth of the questions to be included in the instrument. Testing and interview were conducted both in the UK and Indonesia to ensure a wide sample of student backgrounds and experiences. The instruments used in the UK were MCQ and short answer tests. When answering the multiple-choice questions, students were asked to explain how they arrived at their answers. Deployment of the instrument in the UK was to 57 first year chemistry students and 80 first year food science students taking a foundation chemistry course. All had previously carried out some fundamental chemistry that included chemical kinetics. In Indonesia 260 first year chemistry students from the State University of Malang (UM) and Haluoleo University (UHO) responded: The instruments used in Indonesia were revised versions of those used for testing and interviewing in the UK. Several students from each institution were selected for interview sessions to confirm some unexpected responses to the tests.

The instruments (MCQ and short answer tests) used in this step were carried out in English for UoR students and in Indonesian for Indonesian students. This was intended to avoid a language bias, particularly among Indonesian students. However, a separate assessment to determine whether Indonesian students' proficiency in English is sufficient to understand the questions in English adequately was also carried out. The result of this assessment determined the language to be used in the instrument for Indonesian students in the next step (developing the prototype FTDICK).

Defining Students' Unscientific Ideas and Developing the Prototype of an FTDICK

From the testing and interview step, the following misunderstandings were uncovered: a lack of awareness of the concept of successive half-lives; the misconception that equal numbers of reactant molecules in a bimolecular reaction leads to a higher reaction rate, confusion over the effect of concentration changes on reaction rates with different orders of reactants; an increase in concentration of a zero-order reactant will increase the reaction rate; the values of the exponents in a rate law are directly deduced from the stoichiometry of the reaction; the lower the temperature the higher the reaction rate; an increase in temperature increases the activation energy and leading to an increase in reaction rate. Many respondents also had difficulty in interpreting the plot that describes the progress of a reaction. In addition, students conflated theories in chemical kinetics with other topics in physical chemistry, for example, chemical equilibrium, and also demonstrated confusion over some chemical terminology.

Following an analysis of the results from the testing and interview stage, a further instrument was developed. The unscientific ideas and misconceptions were incorporated as incorrect reason options in the FTDICK instrument along with alternative concepts already reported in the literature. In order to construct the questions, various sources were exploited including textbooks and exam papers. All questions were original and developed by the authors. The prototype questions were again scrutinised by physical chemists responsible for teaching chemical kinetics to final year undergraduate level and a small number of the questions adjusted as a result.

The prototype FTDICK was in English because the assessment previously mentioned demonstrated that the Indonesian students' proficiency in English was good enough to allow the study to be carried out in the English language. This is consistent with one of the goals of the Indonesian educational system which is to promote English skills among Indonesian students in order to prepare them for competing in the global market in the future.

Validating the Prototype FTDICK

Validation of the initial instrument involved 271 students from two Indonesian Universities, including 220 students from the State University of Malang and 51 students from Haluoleo University. The instrument used in this step is a prototype FTDICK instrument (Appendix A) as developed in the previous step. It consists of 20 MCQ each with attached reason options. The time allocated for students to carry out the FTDICK test was 1.5 h. Students in their second semester participated in the study. In both Indonesian universities, a basic chemistry module is provided in the first year across two semesters. Basic Chemistry 1 is provided in the first semester and Basic Chemistry 2 in the second semester. All students at both universities who participated in this research had studied the topic of chemical kinetics in the Basic Chemistry 2 module.

Treatment of Data

Each student response for each question and reason was entered into an Excel spreadsheet along with the student's confidence rating for each question and reason. The responses were graded according to the following procedure. A correct answer in the answer tier (A tier) was scored '1' and an incorrect answer was scored '0'. A correct reason in the reason tier (R tier) was scored '1', and an incorrect reason was scored '0'. If both A tier and R tiers were correct, a score of '1' was allocated for both tiers (B tier). If either or both A tier and R tier were incorrect, a score of '0' was allocated to the B tier.

The average value of all students' confidence ratings was calculated for both the answer tier (CR(TA)) and the reason tier (CR(TR)) for each individual response. For example, CR(TA) = B 2.5 means that the average of all students' confidence ratings who selected option B as their answer in the answer (A) tier is 2.5. CR(TR) = C 2.5 means that the average of all students' confidence ratings who selected option C as their reason in the reason (R) tier is 2.5. The average value of the students' average confidence rating in both their answer (CR(TA)) and their reason (CR(TR)) was also calculated and defined as CR(TB), i.e., the average confidence rating for both tiers.

Several terms and parameters were used to determine the level of students' misconceptions based on the students' confidence ratings as adopted from the previously published study [23]. These terms relate to answer or reason or combined answer and reason calculations that were graded as zero (i.e. CR(TA) of wrong answer, CR(TR) of wrong reason, CR(TB) of wrong answer-correct reason, CR(TB) of correct answer-wrong reason, and CR(TB) of wrong answer-wrong reason).

The terms used are: a genuine misconception which is a misconception expressed with a confidence rating of 2.75 or higher: this can be ascribed due to knowledge gaps and/or wrong reasoning/unscientific ideas. A spurious misconception is a misconception that is expressed with a confidence rating lower than 2.75: it can be argued that this is due to guessing. Genuine misconceptions are further categorized into moderate misconceptions (those expressed with medium level confidence ratings - between 2.75 and 4.0) and strong misconceptions (those expressed with high confidence ratings of 4.0 and above).

Other variables were also calculated in order to give supportive information in determining the level of students' understanding. The variables are described below. Overall mean confidence (CF) which is calculated based on the total of students' confidence ratings divided by the total number of students. The confidence of students giving the correct answers (CFC) which is calculated based on the average of students' confidence ratings who gave the correct answer. The confidence of students when they gave wrong answers (CFW) which is calculated based on the average of students' confidence ratings who gave the wrong answer; Confidence discrimination quotient (CDQ = CFC-CFW/standard deviation of confidence). The CDQ indicates whether the participants can discriminate between what they know and what they do not know. These parameters are calculated for every tier and every question. This method of handling the results is in line with previous research carried out in psychology [33,35] and with studies using a similar four-tier instrument in physics and chemistry [26,28-29].

Item Analysis

Before being used for data collection, the instrument was reviewed for clarity of language, and scientific accuracy by staff in the Department of Chemistry and student feedback was secured to ensure the instrument was understandable.

After data collection, the instrument was analyzed in terms of the difficulty level (DL), discriminatory index (DI), distractor effectiveness, validity and reliability (Table 1). The DL refers to the proportion of students who answer the question correctly. This parameter is determined by dividing the number of students who answer correctly by the total number of students. The DI represents the ability of an item to differentiate between students on the basis of how well they understand the concept being tested. This parameter is determined by subtracting the proportion of students who answer the question correctly from the upper group (those with an above average score on the questions) and the proportion of students who answer the question correctly from the lower group (those with a below average score on the questions). Another useful parameter in reviewing the effectiveness of a test item is the distractor effectiveness. All of the incorrect options, or distractors, should be chosen by at least one student for the option to be effective. Validity refers to whether the information obtained from a test represents the actual understanding of students. Reliability is the degree to which a test consistently measures whatever it is designed to measure. Validity and reliability are measured by the SPSS program using product moment correlation and Cronbach alpha [50] respectively.

RESULTS AND DISCUSSION

Collecting Students' Unscientific Ideas

A total of 397 first-year chemistry students responded to the initial questionnaire survey, and 21 students (3 UK and 18 Indonesian students) took part in subsequent interviews. The small number of students participating in interviews is due to the following factors. Firstly, students invited to attend follow-up interviews were only those who provided some unclear answers in the initial test. Generally, students' answers to the question were clear and easily interpreted by the authors. Secondly, some invited students declined to participate in an interview. A pattern of unscientific ideas and

Table 1. The criteria used to interpret the item analysis parameters

Difficulty level (DL)		Discrimination index (DI)		Reliability		Validity	
value	criteria	value	criteria	value	criteria	$r_{xy\text{-}tab} = 0.113$	criteria
0.00 – 0.30	Hard	0.00 - < 0.10	poor	> 0.90	excellent	$R_{xy} > r_{xy\text{-}tab}$	valid
0.31 – 0.70	Moderate	0.10 - <0.30	fair	0.81 - 0.90	very good	$R_{xy} < r_{xy\text{-}tab}$	invalid
0.71 – 1.00	Easy	0.30 - < 0.75	good	0.61- 0.80	good		
		0.75 – 1.00	excellent	0.40 - 0.60	fair		
		negative	unsuitable	< 0.4	poor		

misunderstandings were collected from the results. The most common of the unscientific ideas are presented below, categorized according to the concept.

Half-life

It appeared that many students only memorize the definition of half-life without an adequate understanding of its conceptual meaning. Some students fully understand that the concentration of a reactant at the half-life, $[A]_{t½}$, is half of the concentration at its initial value, $[A]_{t=0}$. However, many are unclear about whether the relationship that the concentration of $[A]_{t½} = ½[A]_{t=0}$ prevails for any reaction order. This finding is strengthened by the small percentage of students who chose the correct answer but failed to provide the scientific reason for their answer. Several students knew that the concentration of a reactant at its half-life is half of its initial concentration, but assumed this is only applicable for a reactant that decays according to first-order kinetics. Many students struggled to differentiate between the first half-life (t½) and the second half-life (t¼) in a reaction. They understood these half-lives to have the same value and to be independent of reaction order. Such phenomena have not previously been reported in the area of chemical kinetics. Another previously published finding in this topic is that when defining half-life, students often associate it with radioactive decay [34].

Relationship between concentration and rate of reaction

Many students reasoned that equal numbers of reactant molecules lead to a higher reaction rate because if the number of molecules of starting reagents X and Y is equal the maximum amount of product will be formed. This misconception is also a new finding in chemical kinetics. In addition, many students were not aware that an increase or decrease in the concentration of a zero-order reactant does not affect the reaction rate. They reasoned that the effect of concentration for both orders (first and zero-orders) on the reaction rate is the same which supports previous results [30].

Some students believed that the slow step in a reaction mechanism is zero-order. These students also thought that the time needed to complete the reaction was actually the reaction rate. Some students believed that a reactant with a lower concentration will always disappear more quickly than one with a higher concentration. Discussions revealed that they confused their understanding of this topic with that of the limiting reagent. Difficulties and inconsistencies in chemical terminology could be the reason for this apparent lack of knowledge.

A common misconception which was found in this topic is that an increase in the concentration of a zero-order reactant will increase the reaction rate. This finding is similar to the one uncovered by [36]. A small fraction of students argued that the increase in the concentration of a zero-order reactant would decrease the reaction rate. One student claimed that an increase in concentration of a zero-order reactant does not affect the rate because it obeys the law of conservation of mass. Students have been reported to have difficulty understanding and explaining the nature of zero-order reactions [34] and frequently use the mathematical expression for the integrated rate law without a full understanding of what this implies. Another unscientific understanding uncovered is that an increase in the volume of the reaction increases the rate. This confirmed similar findings published by [37] and [38]. They found that several students believed that the change of various factors disturbing the system at equilibrium (increasing concentration of product, increasing temperature, and decreasing volume of a system) increases the rate of the forward reaction, but decreases the reverse one. In addition, a small proportion of students explained that an increase in the concentration of a first-order reactant would increase the rate more than an increase in concentration of a second-order reactant.

The rate law

The most common misconception found within this topic is that the values of the exponents in the rate law are directly deduced from the stoichiometry of the reaction. This confirms previously published results [30,39-40]. In addition, some students struggled to identify the order of a reactant from the rate law provided. Up until now, there has been little published in the literature to explain students' misconceptions regarding the rate law. However, the way in which some popular textbooks introduce the concept of the rate law has provided some interesting results [41], and this study could be used as a starting point for further exploration of the topic.

Effect of temperature on rate of reaction

A few students claimed that temperature does not affect the reaction rate. This misconception is in line with the previous finding [36,41]. Surprisingly several students stated that the lower the temperature, the higher the reaction rate. Another misconception found is that an increase in temperature increases the activation energy and this leads to an increase in reaction rate. This misconception is also in keeping with the previous findings [41-42]. One incomprehensible idea uncovered in this study is that the higher the temperature, the stronger the attractive forces between different molecules.

Collision Theory and Activation Energy

The results showed significant confusion when students used collision theory to describe a reaction on the microscopic scale and also revealed that many students experience difficulty when interpreting diagrams describing the progress of a reaction. Much of the work carried out so far on this topic has concerned students' understanding of the definition of activation energy.

Students were found to have a limited understanding of the concept of collision theory and its effect on reaction rate as evidenced by the vast majority of students who failed to answer questions on this topic. Of those who did answer or were interviewed several unscientific ideas were uncovered. Although most students understood that temperature affects the rate of a reaction they were not aware of the importance of orientation of colliding molecules on the success of a collision and assumed all collisions result in a reaction. Another unscientific idea revealed in this topic is that the collision will be more effective if the colliding particles have a large charge difference. They argued that such a collision would have more impact and therefore lead to an increase in rate. These misunderstandings have not been published so far in the literature. One of the most serious misconceptions is that the lower the temperature, the more effective the collision which leads to a higher rate of reaction. This agrees with the previous finding [43] who found a number of high school students believed that an increase in the initial temperature of the system decreases the reaction rate. They reasoned that collisions between fast-moving particles would be less effective because the particles would be more likely to rebound.

Conflation of theories

Many unscientific ideas revealed can be explained by students' conflating the theory of chemical reaction kinetics with other topics in physical chemistry. Rarely are individual chemical theories unrelated to other areas of chemistry but it is well known that students inappropriately confuse kinetics with equilibria [23,43-44], for example, and thermodynamics with kinetics confusing ideas of thermal stability and kinetic lability [36,45].

Some examples of this were found in the results collected, such as the idea that increasing the temperature will increase the rate of an exothermic reaction. One rationale used by students to explain this is that if the reaction is exothermic in the forward direction an increase in temperature will increase the concentration of product molecules. Two points can be drawn from this reasoning. The first is that students are confusing the idea of reaction rates and Le Châtelier's principle as revealed previously [41]. Secondly, the students are actually applying Le Châtelier's principle incorrectly because an increase in temperature favors the endothermic direction. Similar unscientific conflations between both concepts were reported [32,41].

A further misconception revealed was that a high concentration of reactant and low concentration of product always results in a higher rate. Students harboring this understanding are likely to have confused chemical kinetics with Le Châtelier's Principle and assumed that altering the equilibrium mixture to produce more products increases the rate.

Finally, a few students considered the reactant with the fewer number of particles or lower concentration as the solute, while that with a higher concentration as the solvent. This possibly derives from a misinterpretation of the pictorial representation of the question where the

Table 2. Difficulty level and discriminatory index for the answer tier (A), the reason tier (R) and the combined tiers (B) of the prototype FTDICK instrument

Que	Difficulty Level (DL)			Discriminatory Index (DI)			Que	Difficulty Level (DL)			Discriminatory Index (DI)		
	A tier	R tier	B tier	A tier	R tier	B tier		A tier	R tier	B tier	A tier	R tier	B tier
1	0.18	0.11	0.01	0.33	0.04	0.04	11	0.39	0.07	0.03	0.45	0.08	0.11
2	0.55	0.61	0.46	0.38	0.36	0.59	12	0.44	0.30	0.17	0.41	0.30	0.30
3	0.26	0.07	0.04	0.04	0.16	0.08	13	0.08	0.16	0.01	0.03	0.12	0.03
4	0.51	0.44	0.32	0.58	0.45	0.66	14	0.25	0.30	0.06	0.16	0.40	0.12
5	0.56	0.46	0.35	0.53	0.44	0.74	15	0.32	0.37	0.14	0.27	0.41	0.23
6	0.16	0.20	0.06	0.22	0.04	0.15	16	0.37	0.28	0.13	0.40	0.33	0.26
7	0.48	0.51	0.26	0.47	0.34	0.47	17	0.28	0.27	0.16	0.47	0.52	0.38
8	0.32	0.36	0.20	0.40	0.48	0.38	18	0.28	0.27	0.12	0.40	0.42	0.30
9	0.51	0.19	0.07	0.19	0.07	0.14	19	0.32	0.19	0.08	0.25	0.08	0.08
10	0.42	0.39	0.17	0.18	0.26	0.05	20	0.17	0.06	0.03	-0.03	0.08	-0.01
Mean								0.33	0.28	0.14	0.30	0.27	0.25

Red = hard category yellow = good; teal = fair
Green = moderate category pink = poor; dark yellow = unsuitable

mixture of two different reactants (represented by different colored spheres) was assumed to be a solution. This point was taken into consideration when constructing the final instrument. The students' explanation to support their answer that "If the concentration of solvent is much higher than the concentration of solute, the reaction rate increases" confirmed students' misconception around this point.

Terminology

Reaction kinetics is a topic that abounds with scientific terms that are frequently confused. For example, the different terms used to describe the rate equation such as rate law, rate expression, the rate of reaction and rate and the lack of precision often encountered in textbooks that use these terms interchangeably. Another area of confusion is reaction order which may be the overall order of the reaction as defined by the sum of the individual orders in the rate equation, or the order concerning a specific reagent.

Our results showed that some students confuse the rate of reaction with a time of reaction by stating that an increase in the concentration of a reactant leads to the time of reaction being longer. This finding is similar to that described previous studies [39,46]. Another misunderstanding which could be due to the confusion of reaction time and rate is that the longer the reaction takes, the faster the rate.

Item Analysis (The Result of Validation of the Prototype FTDICK)

The majority of the unscientific ideas and alternative concepts described above were incorporated as incorrect answers and reasons in the prototype FTDICK. In addition, unscientific ideas from the literature and the authors' experiences of teaching chemical kinetics were also drawn upon. The instrument was reviewed for clarity of language, and scientific accuracy by staff in the Department of Chemistry and student feedback was secured to ensure that the language of the instrument was clear. Both chemistry staff and students were given a form to assess the instrument. The instrument was assessed in term of 3 descriptors including language, concept level, and difficulty level. Their feedback was considered and used to revise and finally to produce the final prototype of FTDICK. The instrument was presented to 271 first year Indonesian chemistry students, and the results are described below. Several parameters were used to describe the quality of the prototype FTDICK instrument, including the difficulty level (DL), the discriminatory index (DI), distractor effectiveness, validity, and reliability.

Table 3. Values of distractor effectiveness (%) of the Prototype FTDICK instrument

Question	1		2		3		4		5		6		7	
Option	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	26.94	25.09	4.80	16.24	34.69	11.44	9.23	19.93	4.80	46.13	28.04	34.69	47.97	22.51
B	10.33	22.88	18.08	61.25	33.95	35.06	50.55	21.03	18.82	9.23	15.87	17.71	8.49	50.55
C	39.85	10.70	54.61	7.01	25.83	20.66	24.35	44.28	55.72	23.25	16.24	19.93	11.81	10.70
D	18.08	18.08	5.17		2.95	13.65	11.81	7.01	15.13	11.44	25.09	14.76	23.25	
E						7.38								
F						1.48								

Question	8		9		10		11		12		13		14	
Option	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	10.70	6.64	26.20	12.18	18.45	11.07	19.93	12.18	19.19	21.77	52.03	43.54	14.39	12.18
B	19.56	24.72	50.55	8.12	13.28	39.11	13.28	15.13	19.56	19.93	9.96	15.87	29.52	19.56
C	25.46	35.79	6.27	18.82	41.70	9.23	18.82	29.15	43.91	11.07	21.40	13.65	20.30	29.89
D	32.10	14.39	10.33	14.02	19.19	12.18	39.48	15.13	5.90	29.89	8.12	10.33	25.09	9.59
E				25.83		7.38		6.64						7.38
F				4.80		2.21		1.48						2.95
G						5.17		4.43						
H						0.74								

Question	15		16		17		18		19		20	
Option	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	10.33	36.53	20.30	19.93	19.19	27.31	24.35	27.31	6.27	25.09	36.90	7.38
B	31.37	22.14	36.53	15.50	28.41	18.82	19.19	19.19	39.85	31.37	18.82	18.82
C	31.73	15.50	12.55	17.71	16.97	30.63	12.92	21.03	31.73	19.19	16.61	19.56
D	18.45	13.65	17.34	27.68	21.40	4.43	28.04	11.07	4.80		11.81	18.82
E												8.12
F												5.90

Difficulty level (DL)

Table 2 gives the DL and DI for each item of the instrument. The average DL of all questions is 0.33 for the A tier, 0.28 for the R tier and 0.14 for B tier. These values indicate that the test was reasonably difficult and only those students with a good understanding of the problem are able to answer the questions correctly and select the correct reason for their answers. The DL value of the reason tier is lower than that of the answer tier. This supports the previous finding [4] that many students answer the A tier better than the R tier because in answering the A tier, students simply apply their content knowledge, but a good conceptual reasoning is required to answer the R tier correctly. Furthermore, the DL of the B tier is always the lowest of all the tiers. This trend implies that only students with a genuine understanding can answer both A and R tiers correctly.

Discriminatory index (DI)

The discriminatory index (DI) represents the ability of an item to differentiate between students on the basis of how well they understand the concept being tested. The higher the DI value, the better the question is in discriminating between high and low achieving students. The DI values of questions are presented in Table 2. DI values of the A tier, R tier and B tier ranged from poor to good, none of them was excellent. 40% of the questions, i.e., Q2, Q4, Q5, Q7, Q8, Q12, Q17, and Q18 are categorized as good in each of the A, R and B tiers. This phenomenon indicates that these questions are reasonably able to discriminate between upper and lower ability students.

Meanwhile, the rest of questions have DI grades which differ between tiers. For instance, Q11 has good and fair ratings for the A and B tiers respectively, but a poor category for the R tier. This phenomenon suggests that the reason options for this question need to be considered and should be revised or replaced.

Q20 has negative indices of DI of -0.03 and -0.01 for the A and B tiers respectively. These indices indicate that the question is unsuitable in discriminating between high and low achieving students. The indices strongly suggest that the question should be revised or substituted. When deciding whether a question should be revised or replaced the values of all parameters should be considered. In some circumstances, even a question with a poor DI can be retained because the primary purpose of this instrument is to identify students' understanding instead of differentiating between high low achieving students [47]. The two parameters (DL & DI) confirm that including both the A and R tiers in the instrument is appropriate to identify the students' understanding of chemical kinetics.

Distractor effectiveness

This parameter indicates whether an incorrect answer or reason option is selected in order to highlight those students with a poorer understanding of the concept. The distractor indices for Q1 option A (Table 3) are 26.94 and 25.09 for the A and R tiers respectively. This means that 26.94% of all students selected answer A and 25.09% selected reason A. All of the answers and reasons were selected by some of the students which imply that they are all functional [48]. In total 91.95% of the answer and reason options were selected by more than 5% of students. The values for the individual tiers are 95% of the answers in the A Tier, and 89.36% of the reasons in the R tier were selected by more than 5% of students.

Validity

Validity refers to whether the information obtained from a test represents the actual understanding of the examinees [49]. The validity is shown by the value of the Pearson correlation index (r_{xy}). To determine whether an item or question is categorized as valid or not valid, the value of $r_{xy\text{-calculation}}$ of each item is compared with the value of $r_{xy\text{-table}}$. The higher the r_{xy} value, the greater the validity which means that students' answers to the question represent their actual understanding.

As shown by Table 4, most of the questions are valid. The only invalid question in the A tier is Q20 which is considered to be a difficult question as shown by the low value of the DL of 0.17. This indicates that even high achieving students struggled to answer the question correctly. The question cannot discriminate well between students with a good understanding and students with a poor understanding as confirmed by the value of the DI of -0.03. The DI value implies that the number of students who answered the question correctly from the lower ability group is slightly higher than from the higher ability group which could be ascribed to guesswork.

The validity of tier B Q20 (0.01) is the lowest of all the validity indices. This low validity is also supported by

Table 4. The validity of each item/question of the prototype FTDICK instrument with a confidence level of 95%

Question		1	2	3	4	5	6	7	8	9	10
A tier	r_{xy}	0.34	0.47	0.13	0.53	0.55	0.28	0.39	0.38	0.33	0.23
	category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
R tier	r_{xy}	0.07	0.41	0.21	0.46	0.45	0.09	0.41	0.47	0.18	0.30
	category	Invalid	Valid	Valid	Valid	Valid	Invalid	Valid	Valid	Valid	Valid
B tier	r_{xy}	0.24	0.48	0.21	0.60	0.58	0.31	0.42	0.46	0.27	0.26
	category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
Question		11	12	13	14	15	16	17	18	19	20
A tier	r_{xy}	0.43	0.44	0.14	0.29	0.29	0.40	0.36	0.38	0.28	0.06
	category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Invalid
R tier	r_{xy}	0.15	0.27	0.14	0.39	0.45	0.39	0.47	0.50	0.17	0.19
	category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid
B tier	r_{xy}	0.35	0.36	0.17	0.20	0.33	0.39	0.50	0.48	0.24	0.01
	category	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Valid	Invalid

Table 5. Reliability of items of the prototype FTDICK with confidence level 95% using Cronbach Alpha

General confidence	Answer (A)		Reason (R)		Both (B)	
	Score	Confidence	Score	Confidence	Score	Confidence
0.78	0.59	0.91	0.55	0.92	0.65	0.96

the negative value of the DI that strengthens the suggestion that guessing plays a role here as found previously [26]. The confidence ratings linked to both tiers provide an important means of distinguishing between genuine misunderstandings and guesswork.

Reliability

Reliability is the degree to which a test consistently measures whatever it sets out to measure. Results for reliability of the instrument, as calculated by SPSS, are depicted in Table 5.

The reliability of both A and R tiers are 0.59 and 0.55 respectively. When both tiers are considered, the reliability is slightly higher at 0.65. This phenomenon is in line with the finding published previously [4]. The reliability of the students' confidence rating is higher in the B tier than in the A or R tiers. It can be argued that students giving correct answers when responding to both the A and R tiers simultaneously implies a good understanding of the concept.

Validity and reliability, in particular, are the most important parameters in determining the quality of an instrument [50] and results here have shown that this prototype FTDICK instrument is valid and reliable and can be used for future purposes. However, the results from the item analysis suggested that several questions should be revised and even replaced (5 parameters). For example, Q3 of the A tier and Q11 of the R tier show low validity with r_{xy} values of 0.13 and 0.15 respectively. These low validities are in agreement with the low values of DI and distractor effectiveness (see Tables 2 and 3). Such observations highlighted the questions to be revised or substituted. A further reliability index provided in Table 5 is the reliability of students' general confidence. This index is intended to explore the relationship between students' confidence rating in chemistry as a whole and students' confidence in answering a particular question. This is a novel aspect not generally included in similar four-tier instruments.

The quality of the revised FTDICK to identify students' understanding in the area

The analysis results of 5 parameters (difficulty level, discriminatory index, distractor effectiveness, validity, and reliability) as discussed above confirm that this instrument is transferable and can be used to identify students' understanding for the wider respondents. In the next step the revised instrument was used in a subsequent study carried out with larger sample numbers and over internationally diverse cohorts (Indonesian and UK students). The conclusions highlighted by the present initial investigation will be fully explored in the next report. This will allow us to confirm whether the instrument is transferable across cohorts and to fully identify areas of concern for chemical kinetics educators.

Confidence Ratings

Students' confidence ratings were used to classify whether students' incorrect answers were due to an alternative concept or misunderstanding, lack of knowledge or guesswork. A detailed discussion of students' unscientific ideas in chemical kinetics as investigated by the revised instrument will be presented in the following paper. A summary of students' confidence ratings when providing wrong answers in the A, R and B tiers is presented in Table 6. The table shows that unscientific ideas are held by many students. From the "popular wrong answer" column, genuine misconceptions can be seen to exist for most of the questions in the A tier except Q5 and Q19 which have spurious misconceptions with values of CR(TA) = 2.45 and 2.67 respectively. A similar trend is seen for the R and B tiers. Students' confidence ratings when giving either an incorrect answer in the A tier or incorrect reason in the R tier, or both an incorrect answer and incorrect reason, are generally lower than the confidence rating for the two tiers combined (CR(TB)). Most CR(TB)

Table 6. Students' confidence ratings of wrong answers

Q	Popular wrong answer									Wrong answer with highest CR								
	A tier			R tier			B tier			A tier			R tier			B tier		
	O	A	%	O	R	%	O	B	%	O	CR	%	O	CR	%	O	CR	%
1	C	3.34	39.9	A	2.79	25.1	CA	3.3	12.55	B	3.79	10.3	C	3.21	10.7	BB	3.9	1.48
2	B	2.84	18.1	A	2.59	16.2	BB	3.3	9.60	B	2.84	18.1	A	2.59	16.2	DB	3.8	1.85
3	A	3.5	34.7	B	2.83	35.1	BB	3.2	12.92	A	3.5	34.7	F	3.0	1.48	AA	3.5	3.69
4	C	3.29	24.4	B	2.98	21.0	BA	3.9	7.38	D	3.5	11.8	B	2.98	21.0	BA	3.9	7.38
5	B	2.45	18.8	C	2.79	23.2	CC	2.7	12.55	A	3.0	4.8	C	2.79	23.2	DC	3.5	1.85
6	A	3.37	28	A	3.28	34.7	AA	3.3	16.97	C	3.5	16.2	A	3.28	34.7	DA	3.8	8.49
7	D	3.38	23.2	A	2.9	22.5	AA	3.3	13.28	D	3.38	23.2	A	2.9	22.5	AA	3.3	13.28
8	C	2.91	25.5	B	2.9	24.7	CB	3.1	11.44	C	2.91	25.5	B	2.9	24.7	DB	3.7	3.32
9	A	3.3	26.2	E	3.24	25.8	BE	3.6	19.19	A	3.3	26.2	F	3.31	4.8	DC	3.9	2.95
10	D	3.38	19.2	D	2.76	12.2	AB	3.3	9.23	D	3.38	19.2	G	3.21	5.17	BD	3.8	1.11
11	A	3.09	19.9	C	2.82	29.2	DC	3.1	11.07	B	3.69	13.3	G	3.92	4.43	BE	4.9	2.21
12	B	3.21	19.6	A	2.81	21.8	CA	2.8	9.23	B	3.21	19.6	A	2.81	21.8	DC	3.9	1.85
13	A	3.43	52.03	A	2.87	43.54	AA	3.3	32.97	C	3.5	21.4	D	2.96	20.3	DA	4.5	0.74
14	B	2.98	29.52	B	3.32	19.56	BC	3	12.92	C	3.16	20.3	B	3.32	19.56	AE	3.6	3.32
15	B	3.18	31.37	B	2.83	22.14	BA	3.2	9.96	B	3.18	31.37	C	2.93	15.5	BC	3.4	4.06
16	A	2.8	20.3	A	2.67	19.93	BC	2	8.12	D	3.17	17.34	B	2.69	15.5	AD	3.4	5.54
17	D	3.07	21.4	C	2.76	30.63	DC	3.1	14.39	C	3.22	16.97	C	2.76	30.63	CA	3.7	4.06
18	A	3.33	24.35	B	2.37	19.19	AA	3.5	8.49	A	3.33	24.35	D	2.9	11.07	AA	3.5	8.49
19	B	2.67	39.85	B	2.38	31.37	BB	2.3	15.87	B	2.67	39.85	A	2.93	25.09	DA	3.3	1.11
20	A	3.1	36.9	C	2.77	19.56	AD	2.8	11.81	A	3.1	36.9	E	3.18	8.12	AF	3.5	1.48

% = the percentages of all students who participate in this study; Q = Question; CR = confidence rating; O = option

values are higher than 3.0, but in some cases, notably Q4 and Q9, the CR(TB) values are as high as 3.9 and 3.6 respectively which implies a high degree of misunderstanding in the relevant concepts. Table 6 shows how students' misunderstandings, lack of knowledge and guesses are categorized according to the confidence ratings for each tier and each item.

In general, the A tier requires students to apply content knowledge while the R tier requires them to use scientific reasoning to justify their answer. Some published studies have stated that the R tier is, therefore, more difficult than the A tier [29]. The DL values as presented in Table 2 support this assertion. Students' confidence ratings in the A and R tiers as depicted in Table 6 confirm the difference between students' confidence levels of the tiers.

CONCLUSION

This study has revealed that the prototype FTDICK instrument is valid and reliable. The difficulty of individual items and the discriminatory effectiveness have been identified and the instrument shown to be of a suitable nature to produce meaningful results about students' misconceptions in chemical kinetics. Following the analysis of the instrument, a small number of items were replaced or revised due to their being too hard to produce valid results or because certain distractors were not effective. The FTDICK developed is a relatively new design with confidence ratings attached to both question and answer tiers. Such four-tier instruments have not yet been thoroughly tested, and so the results presented here are timely. The results from this study confirm that including the confidence rating in the A and R tiers of the FDICK instrument provides an effective manner by which to categorize students' misunderstandings. By considering students' confidence ratings, certain incorrect answers could be ascribed due to lack of knowledge or guesswork rather than to misconceptions.

ACKNOWLEDGMENTS

We would like to thank the Directorate General of Higher Education, Ministry of Research, Technology and Higher Education, the Republic of Indonesia for a Ph.D. scholarship to one of us (Habiddin).

REFERENCES

[1] Hailikari, T.K., and Nevgi, A., 2010, How to diagnose at-risk students in chemistry: The case of prior knowledge assessment, *Int. J. Sci. Educ.*, 32, 2079–2095.

[2] Seery, M.K., 2009, The role of prior knowledge and student aptitude in undergraduate performance in chemistry: a correlation-prediction study, *Chem. Educ. Res. Pract.*, 10 (3), 227–232.

[3] Bennett, J.M., and Sozbilir, M., 2007, A study of Turkish chemistry undergraduates' understandings of entropy, *J. Chem. Educ.*, 84 (7), 1204–1208.

[4] Caleon, I., and Subramaniam, R., 2010, Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves, *Int. J. Sci. Educ.*, 32 (7), 939–961.

[5] Carson, E.M., and Watson, J.R., 2002, Undergraduate students' understanding of entropy and gibbs free energy, *Univ. Chem. Educ.*, 6, 4–12.

[6] Treagust, D.F., 2002, "Teaching and Learning About Chemical Compounds" in *Chemical Education: Towards Research-based Practice*, Eds., Gilbert, J.K., De Jong, O., Justi, R., Treagust, D.F., and Van Driel, J.H., Springer, Netherlands, 187–188.

[7] Taştan, Ö., Yalçinkaya, E., and Boz, Y., 2010, Pre-service chemistry teachers' idea about reaction mechanism, *J. Turk. Sci. Educ.*, 7 (1), 47–60.

[8] Habiddin, 2018, Development of a Four-Tier Diagnostic Instrument in Chemical Kinetics (FTDICK) to Investigate First-Year Students' Understanding and Misconceptions in the Area, *Thesis*, University of Reading, United Kingdom.

[9] Osborne, R.., and Gilbert, J.K., 1980, A method for investigating concept understanding in science, *Eur. J. Sci. Educ.*, 2 (3), 311–321.

[10] Novak, J.D., 1990, Concept mapping: A useful tool for science-education, *J. Res. Sci. Teach.*, 27 (10), 937–949.

[11] Peterson, R.F., Treagust, D.F., and 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, *J. Res. Sci. Teach.*, 26 (4), 301–314.

[12] Taber, K.S., 1999, Ideas about ionisation energy: A diagnostic instrument, *Sch. Sci. Rev.*, 81 (295), 97–104.

[13] Kinchin, I.M., 2000, Using concept maps to reveal understanding: A two-tier analysis, *Sch. Sci. Rev.*, 81 (296), 41–46.

[14] Chandrasegaran, A.L., Treagust, D.F., and Mocerino, M., 2007, The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation, *Chem. Educ. Res. Pract.*, 8 (3), 293–307.

[15] Towns, M.H., and Robinson, W.R., 1993, Student use of test-wiseness strategies in solving multiple-choice chemistry examinations, *J. Res. Sci. Teach.*, 30 (7), 709–722.

[16] Peşman, H., and Eryilmaz, A., 2010, Development of a three-tier test to assess misconceptions about simple electric circuits, *J. Educ. Res.*, 103 (3), 208–222.

[17] Dindar, A.C., and Geban, O., 2011, Development of a three-tier test to assess high school students' understanding of acids and bases, *Procedia Soc. Behav. Sci.*, 15, 600–604.

[18] Treagust, D.F., 1988, Development and use of diagnostic tests to evaluate students' misconceptions in science, *Int. J. Sci. Educ.*, 10 (2), 159–169.

[19] Tan, K.C.D., Goh, N.K., Chia, L.S., and 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, *J. Res. Sci. Teach.*, 39 (4), 283–301.
[20] Tan, K.C.D., and Treagust, D.F., 1999, Evaluating students' understanding of chemical bonding, *Sch. Sci. Rev.*, 81 (294), 75–83.
[21] Tyson, L., Treagust, D.F., and Bucat, R.B., 1999, The complexity of teaching and learning chemical equilibrium, *J. Chem. Educ.*, 76 (4), 554–558.
[22] Hasan, S., Bagayoko, D., and Kelley, E.L., 1999, Misconceptions and the Certainty of Response Index (CRI), *Phys. Educ.*, 34 (5), 294–299.
[23] Voska, K.W., and Heikkinen, H.W., 2000, Identification and analysis of student conceptions used to solve chemical equilibrium problems, *J. Res. Sci. Teach.*, 37 (2), 160–176.
[24] Arslan, H.O., Cigdemoglu, C., and Moseley, C., 2012, A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain, *Int. J. Sci. Educ.*, 34 (11), 1667–1686.
[25] Loh, A.S.L., Subramaniam, R., and Tan, K.C.D., 2014, Exploring students' understanding of electrochemical cells using an enhanced two-tier diagnostic instrument, *Res. Sci. Technol. Educ.*, 32 (3), 229–250.
[26] Sreenivasulu, B., and Subramaniam, R., 2013, University students' understanding of chemical thermodynamics, *Int. J. Sci. Educ.*, 35 (4), 601–635.
[27] Bain, K., and Towns, M.H., 2016, A review of research on the teaching and learning of chemical kinetics, *Chem. Educ. Res. Pract.*, 17 (2), 246–262.
[28] Caleon, I., and 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, *Res. Sci. Educ.*, 40 (3), 313–337.
[29] Sreenivasulu, B., and Subramaniam, R., 2014, Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures, *Res. Sci. Educ.*, 44 (6), 801–828.
[30] Cakmakci, G., and Aydogdu, C., 2011, Designing and evaluating an evidence-informed instruction in chemical kinetics, *Chem. Educ. Res. Pract.*, 12 (1), 15–28.
[31] Yan, Y.K., and Subramaniam, R., 2016, Diagnostic appraisal of grade 12 students' understanding of reaction kinetics, *Chem. Educ. Res. Pract.*, 17 (4), 1114–1126.
[32] Kolomuç, A., and Tekin, S., 2011, Chemistry teachers' misconception concerning concept of chemical reaction rate, *Eurasian J. Phys. Chem. Educ.*, 3 (2), 84–101.
[33] Lundeberg, M.A., Lundeberg, M.A., Fox, P.W., Brown, A.C., and Elbedour, S., 2000, Cultural influences on confidence: Country and gender, *J. Educ. Psychol.*, 92 (1), 152–159.
[34] Bain, K., Rodriguez, J.M.G., and Towns, M.H., 2018, Zero-order chemical kinetics as a context to investigate student understanding of catalysts and half-life, *J. Chem. Educ.*, 95 (5), 716–725.
[35] Stankov, L., and Crawford, J.D., 1997, Self-confidence and performance on tests of cognitive abilities, *Intelligence*, 25 (2), 93–109.
[36] Cakmakci, G., 2010, Identifying alternative conceptions of chemical kinetics among secondary school and undergraduate students in Turkey, *J. Chem. Educ.*, 87 (4), 449–455.
[37] Hackling, M.W., and Garnett, P.J., 1985, Misconceptions of chemical equilibrium, *Eur. J. Sci. Educ.*, 7 (2), 205–214.
[38] Bilgin, I., and Geban, O., 2006, The effect of cooperative learning approach based on conceptual change condition on students' understanding of chemical equilibrium concepts, *J. Sci. Educ. Technol.*, 15 (1), 31–46.
[39] Cakmakci, G., Leach, J., and Donnelly, J., 2006, Students' ideas about reaction rate and its relationship with concentration or pressure, *Int. J. Sci. Educ.*, 28 (15), 1795–1815.
[40] Turányi, T., and Tóth, Z., 2013, Hungarian university students' misunderstandings in thermodynamics and chemical kinetics, *Chem. Educ. Res. Pract.*, 14 (1), 105–116.
[41] Yalçınkaya, E., Taştan-Kırık, Ö., Boz, Y., and Yıldıran, D., 2012, Is case-based learning an

effective teaching strategy to challenge students' alternative conceptions regarding chemical kinetics?, *Res. Sci. Technol. Educ.*, 30 (2), 151–172.
[42] Kingir, S., and Geban, O., 2012, Effect of conceptual change approach on students' understanding of reaction rate concepts, *Hacettepe Univ. J. Educ.*, 43, 306–317.
[43] Van Driel, J.H., 2002, Students' corpuscular conceptions in the context of chemical equilibrium and chemical kinetics, *Chem. Educ. Res. Pract.*, 3 (2), 201–213.
[44] Kousathana, M., and Tsaparlis, G., 2002, Students' errors in solving numerical chemical equilibrium problems, *Chem. Educ. Res. Pract.*, 3 (1), 5–17.
[45] Holme, T., Luxford, C., and Murphy, K., 2015, Updating the general chemistry anchoring concepts content map, *J. Chem. Educ.*, 92 (6), 1115–1116.
[46] Bektasli, B., and Cakmakci, G., 2011, Consistency of students' ideas about the concept of rate across different contexts, *Educ. Sci.*, 36 (162), 273–287.
[47] Suruchi, S., and Rana, S.S., 2014, Test item analysis and relationship between difficulty level and discrimination index of test items in an achievement test in biology, *PIJR*, 3 (6), 56–58.
[48] DiBattista, D., and Kurzawa, L., 2011, Examination of the quality of multiple-choice items on classroom tests, CJSoTL, 2 (2), 4.
[49] Airasian, P.W., 1994, *Classroom Assessment*, Eds., Arkers, L., and Rosenberg, E., Mcgraw-Hill, Singapore.
[50] Kimberlin, C.L., and Winterstein, A.G., 2008, Validity and reliability of measurement instruments used in research, *Am. J. Health Syst. Pharm.*, 65 (23), 2276–2284.