



Abstract. *Student understanding of chemical equilibrium in aqueous solutions (CEAS) plays a vital role in their upper secondary school chemistry learning and everyday life. Diagnosis of students' alternative conceptions (ACs) of the CEAS will provide teachers with valuable information to make instructional decisions on student learning. This study aims to develop and validate an instrument to diagnose students' ACs about the CEAS, including ionization equilibrium, water self-ionization equilibrium, the equilibrium of salt hydrolysis, and precipitation and dissolution equilibrium. Using Treagust (1998)'s development framework, we have developed 25 two-tier multiple-choice items for the CEAS diagnostic test. After completing the corresponding courses, 750 Grade 11 students from five public schools responded to the CEAS diagnostic test. Rasch modeling approach was employed to provide psychometric properties of the CEAS diagnostic test consisting of one-dimensionality, reliability, and validity. This study identified 15 ACs toward the CEAS. This study found that most students performed better on concept tiers rather than reasoning tiers. In addition, students have difficulties in connecting acidity, solubility, ionization, and chemical reaction and in using mathematical thinking to do transformation between concentration, equilibrium constant, and pH value.*

Keywords: *chemical equilibrium in aqueous solutions, alternative conceptions, two-tier multiple-choice items, Rasch modeling*

Peng He

Northeast Normal University, China
Michigan State University, USA

Changlong Zheng

Northeast Normal University, China

Tingting Li

Northeast Normal University, China
Michigan State University, USA



UPPER SECONDARY SCHOOL STUDENTS' CONCEPTIONS OF CHEMICAL EQUILIBRIUM IN AQUEOUS SOLUTIONS: DEVELOPMENT AND VALIDATION OF A TWO-TIER DIAGNOSTIC INSTRUMENT

**Peng He,
Changlong Zheng,
Tingting Li**

Introduction

Chemical equilibrium in aqueous solutions (hereafter CEAS) is elaborated as the state in which all particles, including ions and molecules in an aqueous solution, are present in concentrations that have no further tendency to change over time. In chemistry, the CEAS is one of the essential concepts in upper secondary school science. The CEAS is identified as a particular case of chemical equilibrium and solution chemistry consisting of acid-base equilibrium, self-ionization of water, hydrolysis of salt solutions, and precipitation and dissolution equilibrium (Ministry of Education, P. R. China [MoE], 2017). In everyday life, the CEAS interacts with various real-world phenomena, such as acid-base balance in human body and the mechanism of soda-acid fire extinguishers. To make sense of those ubiquitous phenomena, students must understand the CEAS, including the behaviors of particles (i.e., molecules and ions) and their interactions at a microscope level in an aqueous solution. Making sense of such complex phenomena also requires considerable knowledge and sophisticated understanding of some basic concepts, including the dissolutions of acids, bases and salts, chemical equilibrium. However, evidence indicates that students often could not fully understand those basic concepts, which leads to many ACs (Karpudewan et al., 2015; Orwat et al., 2017; Özmen, 2008). Those ACs are inconsistent with scientifically acceptable ideas, which might be held for students even if they have completed the required courses of chemistry (Palmer 1999). Moreover, ACs about the CEAS make it difficult for students to master the advanced concepts that are built upon those basic concepts (e.g., Calik, 2005). It is also a challenge for students to incorporate several related concepts to understand complex phenomena (Garnett et al., 1995).

To this point, there is an urgent need to diagnose students' ACs about the CEAS that identify students' problems, which could provide useful information to support them better understanding. In addition, understanding students' ACs toward the CEAS can also provide valuable information for teachers to make their instructional decisions (Licht, 1991; Mdachi, 2012). Many existing AC studies have investigated on several related concepts such as chemical equilibrium and solution chemistry (Bilgin & Geban, 2006; Calik, 2005; Demircioğlu et al., 2005; Piquette & Heikkinen, 2005; Voska & Heikkinen, 2000). However, those studies have been conducted at least two decades. The ACs in those studies might not be updated and suitable for now. Some recent studies have also explored the ACs about related to CEAS such as chemistry equilibrium (Akkus et al., 2011), solution chemistry (Adadan & Savasci, 2012), acid-base concepts (Damanhuri et al., 2016), electrolyte concepts (Lu & Bi, 2016), and salt hydrolysis (Orwat et al., 2017). However, they only address some aspects of the CEAS. Little effort has explored students' holistic and in-depth understanding of the CEAS, including acid-base equilibrium, self-ionization of water, hydrolysis of salt solutions, and precipitation and dissolution equilibrium. Notably, students who have a comprehensive understanding of the CEAS can explain the real-world phenomena or solve problems. It is impossible to be accomplished by just employing fragmented and isolated concepts (Odden & Russ, 2019; Songer & Linn, 1999). To achieve this goal, a reliable and valid diagnostic tool needs to be developed and used to assess students' comprehensive conceptions of the CEAS, which can be used for formative and summative purposes.

Diagnostic assessments are commonly used to explore students' ACs in science education. Previous studies have used three types of diagnostic assessments: multiple-choice paper-pencil tests (Artdej et al., 2010; Damanhuri et al., 2016; Demircioğlu et al., 2005), open-ended questions (Kousathana et al., 2005), and interviews (Orwat et al., 2017). Two-tier multiple-choice diagnostic assessments have been widely adopted and used in the past several decades. Treagust's (1988) research-based design process has been highlighted as a robust technique for developing two-tier multiple-choice diagnostic instruments. The developmental procedure consists of three stages: defining the content area, obtaining information about students' ACs, and developing two-tier diagnostic instruments. So far, the framework has been employed and validated by researchers from many countries (e.g., Adadan & Savasci, 2012; Lu & Bi, 2016; Tan et al., 2002). This study employs the framework (Treagust, 1988) to develop a two-tier diagnostic instrument for assessing students' ACs towards chemical equilibrium in aqueous solutions (CEAS). Two measurement theories are commonly employed to validate the quality of two-tier multiple-choice diagnostic instruments: Classical test theory (CTT) and item response theory (IRT). Many two-tier multiple-choice instruments (e.g., Artdej et al., 2010; Damanhuri et al., 2016; Demircioğlu et al., 2005) have been validated by the CTT method. The CTT-based approach provides Cronbach's alpha reliability, difficulty indices, and discrimination indices to validate the diagnostic instruments. However, evidence indicates that the IRT is superior to CTT (Reise & Haviland, 2005). The CTT-based approach uses a common estimate of measurement precision that is assumed to be equal for all individuals irrespective of their attribute levels, while the measurement precision of IRT depends on the latent-attribute value. IRT-based techniques, such as Rasch modeling, are more theoretically grounded and model the distribution of students' abilities at the item level. The IRT-based techniques avoid the problems of sample dependence of item difficulty and item discrimination generated by the CTT-based approach (Morales, 2009). To date, only a few two-tier multiple-choice instruments have been validated by the IRT-based approach (Fulmer et al., 2015; Lu & Bi, 2016). This study employs the Rasch modeling approach to validate the instrument for assessing students' ACs of the CEAS.

Research Aim and Research Questions

This study explores Grade 11 students' ACs of the CEAS in Chinese upper secondary school chemistry classrooms. A two-tier multiple-choice diagnostic test is developed and used to explore students' ACs of the CEAS. The research questions are listed as follows:

- (1) What empirical evidence supports the diagnostic instrument's unidimensionality, reliability, and validity for assessing students' understanding of chemical equilibrium in aqueous solutions? What is the evidence for suggesting further improvements of the diagnostic instrument?
- (2) What are students' alternative conceptions of chemical equilibrium in aqueous solutions in the Chinese upper secondary school chemistry classrooms?



Research Methodology

General Background

As the CEAS is the one of the most challenging topics in upper secondary school chemistry, diagnosing and understanding students' ACs toward the topic is needed. To date, a few studies reported some ACs related to this topic (Artdej et al., 2010; Damanhuri et al., 2016; Demircioğlu et al., 2005), there is no comprehensive diagnostic test specifically designed for this purpose. This study employed Treagust (1988)'s framework to develop a two-tier multiply choice test and assessed Grade 11 Chinese students' understanding of the CEAS. Rasch analysis was used to generate the validity and reliability of the developed test. Students' major ACs were identified and reported in this study.

Participants

Participants were Grade 11 students from three districts (Jilin, Liaoning, and Hebei provinces) in north part of mainland China. Given the constraints of funding and resources, this study employed convenience sampling strategy to recruit students from our collaborative teachers. Compared to those key secondary schools with a great number of high-perform students, ordinary secondary schools were purposely selected to represent a wide range of students' achievements in those three districts. In total, 750 students from five schools were included as our final sample. Of the sample, 212 students in four classes were selected from two schools in Jilin province, 199 students in four classes came from two schools in Liaoning, and the other 339 students in 6 classes were from one school in Hebei province. 396 (52.8%) were males and 346 (46.1%) were females. Eight (1.1%) students didn't report their gender information.

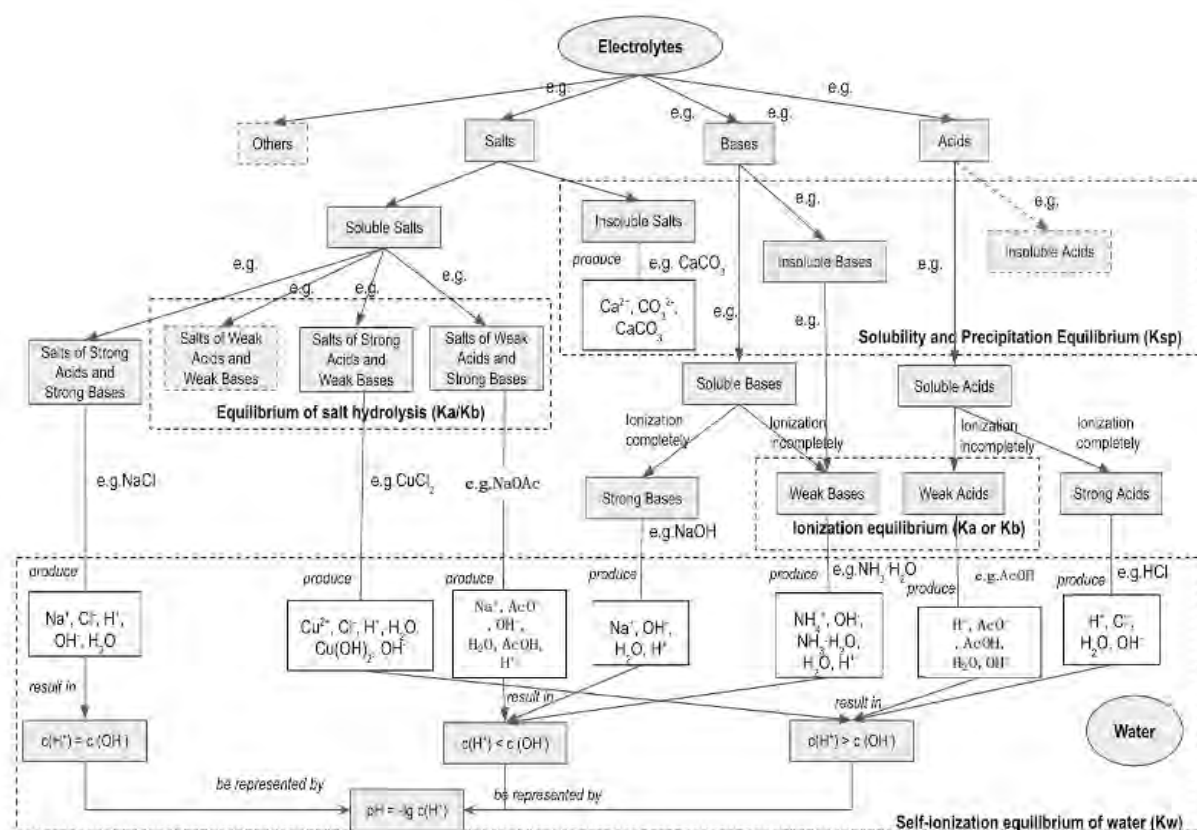
Instrument and Procedures

Based on the Treagust' Framework (1988), the three stages were included in the developmental process to create a two-tier multiple-choice diagnostic instrument named Chemical Equilibrium Aqueous Solutions Test (hereafter CEAS-T).

Defining content area. According to the Chinese Upper Secondary School Chemistry Curriculum Standards (MoE, 2017), four essential concepts of the CEAS were selected from Theme 3: ionization equilibrium of weak electrolytes, the self-ionization equilibrium of water, the equilibrium of salt hydrolysis, and precipitation and dissolution equilibrium of insoluble electrolytes (see Appendix A). Figure 1 presents a concept map related to the propositional statements in the same topic area. Two chemistry education researchers were recruited to review the concept map and the associated propositional statements. They provided feedback on the alignment, coverage, and potential gaps between those two parts. They also pointed out the inappropriate expressions in the propositional statements. Two expert chemistry teachers were further recruited to review the two documents based on their teaching experience. They revised some expression, wording, and language issues to make the items more appropriate for students.



Figure 1
The CEAS Concept Map



Note: "e.g." represents "for example"; "AcOH" represents hydrogen acetate (CH_3COOH), "AcO" represents CH_3COO^- , and NaOAc represents CH_3COONa .

Identifying students' conceptions. Students' conceptions of the CEAS were identified from multiple data sources, including classroom observation field notes, students' homework, and feedback from teachers and students. Appendix B presents an example of student homework using Orwat et al.'s (2017) design. In this example, when students completed the course related to salt hydrolysis equilibrium, their teacher assigned the homework about salts dissolving in pure water. Research assistants helped the teacher in collecting, scoring students' responses, and generating a report for the teacher about their students' major ACs. Next, the teacher and students reviewed and discussed incorrect answers in the exercise lesson while our research assistants observed the classroom. After class, we interviewed three students further to identify their ACs with the teacher's assistance. Lastly, the teacher provided detailed feedback on students' ACs from their homework and helped us confirm the ACs.

Developing items. Consistent with previous studies (e.g., Adadan & Savasci, 2012; Artdej et al., 2010; Demircioğlu et al., 2005; Lu & Bi, 2016), this study developed 25 two-tier multiple-choice items. In each item, the first tier was about students' understanding of the concepts and the second tier was to provide reasons why the choice in the first tier is correct. The correct choices were developed aligned with the propositional statements. The distractor choices were selected from multiple sources including classroom observation field notes, students' homework, and teacher feedback. Table 1 presents the numbers of items in each CEAS concept and shows the examples of propositional statements and associated potential ACs for each exemplar item.

Table 1*The Summary of Items, Exemplar Propositional Statements, and Associated Alternative Conceptions*

| CEAS Concepts | Code (N) | Propositional statements | Potential alternative conceptions (ACs) |
|---|------------|---|---|
| Ionization Equilibrium (IE) | IE1-7 (7) | The equilibrium state of weak electrolyte solutions can be changed by changing temperature, diluting, or adding more solutes, or reacting with other substances. (IE6) | a. Increasing temperature of a weak acid solution does not change the concentration of ions (e.g., H ⁺) in the solution. b. Diluting the weak acid solution can reduce the concentration of ions (e.g., H ⁺) and the number of ions (e.g., H ⁺). c. Adding a salt (sodium acetate) solid into a weak acid (acetic acid) solution increases the concentration of ions (e.g., H ⁺). |
| Self-ionization Equilibrium of Water (SEW) | SEW1-6 (6) | Temperature affects the self-ionization equilibrium of water. As temperature increases, the self-ionization equilibrium of water increases, so the Kw value increases, and the pH value decreases. (SEW4) | a. Increasing temperature does not affect the self-ionization equilibrium of water, so the Kw value and pH value do not change. b. As temperature increases, the self-ionization equilibrium of water increases, so the Kw increases, and the pH value of water increases. c. As temperature increases, the self-ionization equilibrium of water decreases, so the Kw decreases, and the pH value of water increases. |
| Equilibrium of Salt Hydrolysis (ESH) | ESH1-8 (8) | Acid salts (or basic salts) can react with water to form acidic (or basic) solutions. (ESH4) | a. All salts react with water to form either acidic or basic solutions. b. All salts can only be dissolved into the water but cannot react with water. c. Acid salts (e.g., NH ₄ Cl) reacting with water can form basic solutions. d. Basic salts (e.g., Na ₂ CO ₃) reacting with water can form acidic solutions. |
| Precipitation and Dissolution Equilibrium (PDE) | PDE1-4 (4) | When the dissolution and precipitation of solute species occur at equal rates, solubility equilibrium is established. (PDE4) | a. Precipitation and dissolution equilibrium exist only in dissolving insoluble substances but not in dissolving soluble substances. b. When the concentration of substances in a solution does not change, precipitation and dissolution equilibrium does not change. c. When the solution evaporates, the dissolution and precipitation equilibrium does not change. |

Figure 2 presents an example of two-tier multiple-choice items in the instrument. Cognitive interviews of six students were conducted to clarify the item language and expression (Padilla & Leighton, 2017). Two chemistry researchers and two expert teachers were recruited to provide the face and content validity of the instrument (McCoach et al., 2013). 25 two-tier multiple-choice items were included to diagnose students' conceptions of chemical equilibrium in aqueous solutions at Grade 11. The psychometric properties of the instrument are presented in the following Result section.

Figure 2*An Example of Items in the CEAS-T Instrument*

Item IE6

(Concept tier) At room temperature, 0.1mol / L HF solution can be continuously diluted with water. Which of the following numbers can keep increasing? (D)

A. $c(H^+)$ B. $K_a(HF)$ C. $c(H^+) \times c(F^-)$ D. $\frac{c(H^+)}{c(HF)}$

$[Note: K_a(HF) = \frac{c(H^+) \times c(F^-)}{c(HF)}]$

(Reason tier) The reason for my answer is: (3)

- Adding water into the HF solution, the position of equilibrium moves to the right, and the ionization degree increases, so the equilibrium constant increases.
- Adding water into the HF solution, the position of equilibrium moves to the right, so $c(H^+)$ always increases.
- The equilibrium constant only relates to temperature, so the equilibrium constant does not change; $c(F^-)$ is getting smaller and smaller, so it always increases.
- Adding water into the HF solution, the equilibrium position moves to the right, $c(H^+)$ increases, $c(F^-)$ increases, so $c(H^+) \times c(F^-)$ always increases.

Data Collection and Analysis

The data was collected at the end of school year in 2018-2019. Before taking the test, all Grade 11 students had completed the learning of the CEAS in Module 1: Principles of Chemical Reactions that is one of the Optional Compulsory Courses. The test was taken 100 minutes. In addition, before learning about the topic of ionic equilibrium, students had learned the concepts of ionization and ionic reactions related to strong electrolytes from the Compulsory Courses at Grade 10 "Theme 2: Common Inorganic Substances and Their Application" (see Appendix A). No local procedures or commitments were required for this study. All students involved in the survey agreed that their responses could be used for research purposes.

To answer RQ1, Rasch modeling approach was applied to explore the psychometric properties of the CEAS-T instrument, including an estimate of students' ability and item difficulty and indices of model-data fit (He et al., 2016; Liu & Boone, 2006). The raw score of students' abilities and item difficulties can be converted into a non-linear transformation using the Rasch model so that the two parameters can be estimated and mapped at the same scale. The formula of the Rasch model can be found in the previous relevant literature (Liu & Boone, 2006). Two basic assumptions of Rasch model were employed to validate multiple-choice two-tier diagnostic instruments (Fulmer et al., 2015; Lu & Bi, 2016). First, as all items on the instrument were designed to diagnose students' understanding of chemical equilibrium in aqueous solutions, the construct of students' performances should be a single latent trait. Thus, the unidimensionality of the instrument was examined and tested by Rasch analysis. The second assumption is the local independence of the items, which means the response for one item is not affected by the responses of the other items. As a two-tier design of items in the instrument, the response of the reason tier relied on the concept tier. Students' ACs were stored in the combinations of options in both first and second tiers. In this study, the combinations of students' responses on first and second tiers were coded as dichotomous data: if both tiers were correct, 1 point was rated; all other combinations were coded as 0 point (Yang et al., 2018; Lu & Bi 2016; Romine et al., 2015). The Winsteps software was used for Rasch measurements.

To answer RQ2, cross-tabulation was used to explore the consistency of students' answers (Tan et al., 2002). Students' ACs were identified by the percentage of the combined tiers' incorrect responses. When the rate of incorrect answers in the combined tiers was higher than 10%, students' ACs were identified in this study (Peterson, 1986). The percentage of ACs were calculated by using different choice combinations of students' responses on each item's concept and reasoning tiers. For item IE6, the combinations of two-tier choices are A (2), showing that the student selected choice A from the first tier and choice 2 from the second tier. Thus, the student with such answers had an AC that adding water into a weak acidic solution promotes the shift of ionic equilibrium and increases the concentration of hydrogen ions.

Research Results

The Rasch analysis was used to analyze the final dataset of 750 students who responded to the 25 items on the instrument. Five items (IE1, IE5, SEW2, SEW3, and ESH6) were deleted due to defective item fit statistics (the outfit MNSQ values larger than 1.30). The responses of 750 students on the remaining 20 items were further analyzed using the same analytic strategy. Detailed findings are shown as follows.

The Reliability and Validity of the CEAS-T Instrument

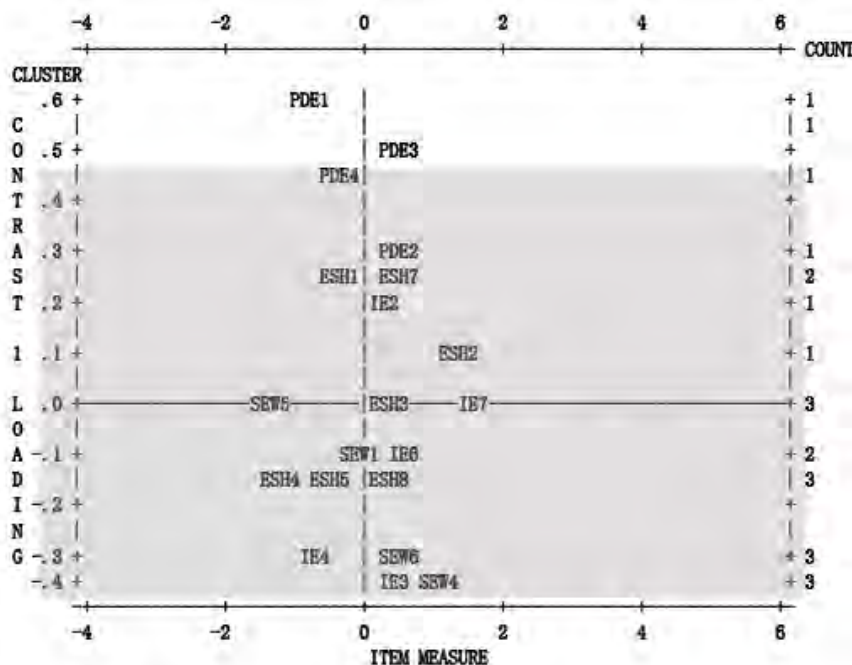
Separation and reliability. In Rasch modeling, the person and item separation and Cronbach's reliability indices were used to show the reliability of the developed instrument. The person separation shows how consistently our estimates of students' abilities match the data, whereas the item separation presents how reliable an instrument can differentiate a wide range of students based on their abilities. In this study, the results show that the person separation index is 2.14, with an equivalent Cronbach's reliability coefficient (alpha value) of 0.82. The item separation index is 6.81, and the corresponding Cronbach's alpha value is 0.98 (see Table 2). The high item reliability indicates that the items of varying difficulties can be differentiated under the model.



Table 2*Summary Statistics of Person and Item*

| Parameter (N) | Measure | INFIT | | OUTFIT | | Separation | Reliability |
|---------------|---------|-------|-------|--------|------|------------|-------------|
| | | MNSQ | ZSTD | MNSQ | ZSTD | | |
| Person (750) | -0.23 | 1.00 | 0.07 | 1.02 | 0.08 | 2.14 | 0.82 |
| Item (20) | 0.00 | 1.00 | -0.11 | 1.02 | 0.11 | 6.81 | 0.98 |

Dimensionality. The principal component analysis (PCA) was used to identify the instrument's dimensionality based on the standardized residuals' variance. However, no detectable residual pattern indicates that all essential variances in the data were modeled by a single Rasch model (Romine et al., 2015). One way could show the plot of item loading graphically; see it in Figure 3. The horizontal axis represents the measured item, and the vertical axis represents the contrast loading between the items and the contrast component. Among the 20 items, the contrast loadings of three items (PDE1, PDE3, and PDE4) are outside the range of -0.4 to +0.4. The more items falling into the range indicate that the instrument is in a unidimensional construct. Another way for detecting the residuals' random noise is to use the eigenvalue of unexplained variance in the first component. The strict criteria of cut-off values are around 2 (Raiche, 2005). Our Rasch analysis shows that the first eigenvalues for PCA on residuals were 1.87, meeting the strict criteria. However, the instrument with 20 items could only explain 31% of the total variance. The result shows that the instrument might still be inadequate to measure students' understanding of CEAS, which was similarly reported in previous studies (e.g., Fulmer et al., 2015; Lu & Bi, 2016). The relatively lower explanatory rate also indicates that the CEAS had complex constructs which were difficult for students to understand.

Figure 3*The Plot of Item Loadings*

The local independence of items was also examined to test the dimensionality of the instrument. Linacre (2009) defined local independence as no commonalities between items accounting for latent construct. To meet the assumption of local independence, the residuals of all items are not expected to be highly correlated with each other. Furthermore, the value of item residual correlations should be smaller than 0.7, indicating less than 50%

shared variance (Linacre, 2013). Our results show that the residual correlations between all items were lower than 0.2, meeting the criteria. The above collective evidence indicates that the dimensionality of the CEAS-T instrument was a unidimensional construct for measuring students' understanding of the CEAS.

Fit statistics. The item fit statistics and person-item estimate map were the two ways to validate the instrument. For the item fit statistics, the mean square residual (MNSQ) and the standardized mean square residual (ZSTD) were used as the fit indices to examine how well each item fits the model. The expected values of MNSQ and ZSTD are 1 and 0. For multiple-choice tests, items have an acceptable fit if their MNSQs fall into the range of 0.7 to 1.3, while ZSTD values are within the scope of -2 to +2 (Wright et al., 1994). Two types of MNSQ and ZSTD are given as Infit and Outfit mean squares in Rasch model. The infit mean squares (MNSQ and ZSTD) are more sensitive to the pattern of responses to items targeted on the person. In contrast, the Outfit mean squares (MNSQ and ZSTD) are more sensitive to responses to items with difficulty far from a person. The Outfit means square is a chi-square sensitive to outliers that often makes lucky guesses for lower ability students or careless mistakes for high ability students. The MNSQ value higher than 1.3 (ZSTD >2) indicates unpredictability that the data underfit the model, whereas the MNSQ value less than 0.7 (ZSTD <-2) indicates too predictability that the data overfit the model (Bond & Fox, 2015).

Based on the above criteria, the results show that most items fitted well with the model, which were identified by the Infit and Outfit values of MNSQs and ZSTDs (See Table 3). Ideally, both the Infits and Outfits of MNSQ and ZSTD should be considered for examining the quality of items. However, for the applicable purpose of the instrument, the MNSQs is a practical data-model fit that is more reasonable than the ZSTDs (perfect data-model fit) and the Outfits are more comfortable to diagnose and remedy than Infits (Linacre, 2013). Therefore, the study found that all 20 items had good model fits. In addition, the point measure correlation (PTMEA) correlates person item scores and person measures (Linacre, 2013). The values of the PTMEA should be positive and not be nearly zero (Bond & Fox, 2015). From Table 3, all items were positive within a range of 0.38 to 0.64.

Table 3
Item Fit Statistics for All Items

| Item | Measure | Error | OUTFIT | | OUTFIT | | PTMEA |
|------|---------|-------|--------|-------|--------|-------|-------|
| | | | MNSQ | ZSTD | MNSQ | ZSTD | |
| IE2 | 0.18 | 0.09 | 0.99 | -0.28 | 1.11 | 1.48 | 0.52 |
| IE3 | -0.18 | 0.09 | 1.09 | 2.17 | 1.12 | 1.84 | 0.50 |
| IE4 | -1.00 | 0.09 | 0.96 | -0.98 | 0.89 | -1.42 | 0.59 |
| IE6 | -0.06 | 0.09 | 1.01 | 0.35 | 1.10 | 1.52 | 0.53 |
| IE7 | 0.95 | 0.09 | 1.05 | 1.14 | 1.14 | 1.37 | 0.44 |
| SEW1 | -1.65 | 0.10 | 0.86 | -2.79 | 0.98 | -0.11 | 0.62 |
| SEW4 | -0.04 | 0.09 | 1.11 | 2.79 | 1.16 | 2.33 | 0.48 |
| SEW5 | -0.39 | 0.09 | 0.96 | -0.90 | 0.93 | -1.02 | 0.57 |
| SEW6 | 0.24 | 0.09 | 0.94 | -1.56 | 0.95 | -0.62 | 0.55 |
| ESH1 | -0.12 | 0.09 | 1.14 | 3.39 | 1.13 | 1.97 | 0.48 |
| ESH2 | 1.07 | 0.09 | 1.10 | 2.21 | 1.15 | 1.35 | 0.42 |
| ESH3 | 0.11 | 0.09 | 1.09 | 2.36 | 1.13 | 1.82 | 0.48 |
| ESH4 | -0.40 | 0.09 | 0.83 | -4.47 | 0.78 | -3.58 | 0.63 |
| ESH5 | -0.38 | 0.09 | 1.07 | 1.83 | 1.10 | 1.47 | 0.52 |
| ESH7 | 0.56 | 0.09 | 0.98 | -0.54 | 1.00 | 0.00 | 0.51 |
| ESH8 | 0.59 | 0.09 | 0.86 | -3.83 | 0.82 | -2.28 | 0.57 |
| PDE1 | -0.66 | 0.09 | 0.89 | -2.62 | 0.76 | -3.85 | 0.62 |
| PDE2 | 0.29 | 0.09 | 1.07 | 1.80 | 1.21 | 2.63 | 0.47 |
| PDE3 | 0.12 | 0.09 | 0.96 | -1.03 | 0.95 | -0.66 | 0.55 |
| PDE4 | -0.15 | 0.09 | 0.79 | -5.91 | 0.76 | -3.96 | 0.64 |



Person-item mapping. Figure 4 shows the persons' ability estimates and the items' difficulty estimates were plotted on the same logit scale in a person-item estimate map. The vertical line on the left is the distribution of students' abilities to understand ionic equilibrium from lower (bottom) to higher (top). The map on the right side is the distribution of items from easy (bottom) to difficult (top) endorsement. A person and an item were at the same position on the logit scale, indicating the person had a 50% probability of answering the item correctly (Bond and Fox, 2015). From the person-item map, this study found that all items were clustered around -2 to +1 on the logit scale with a wide range of students' abilities. However, there were still many students with low abilities (below -2 logit) who might have had a majority of ACs toward the CEAS.

Figure 4
The Person-item Mapping

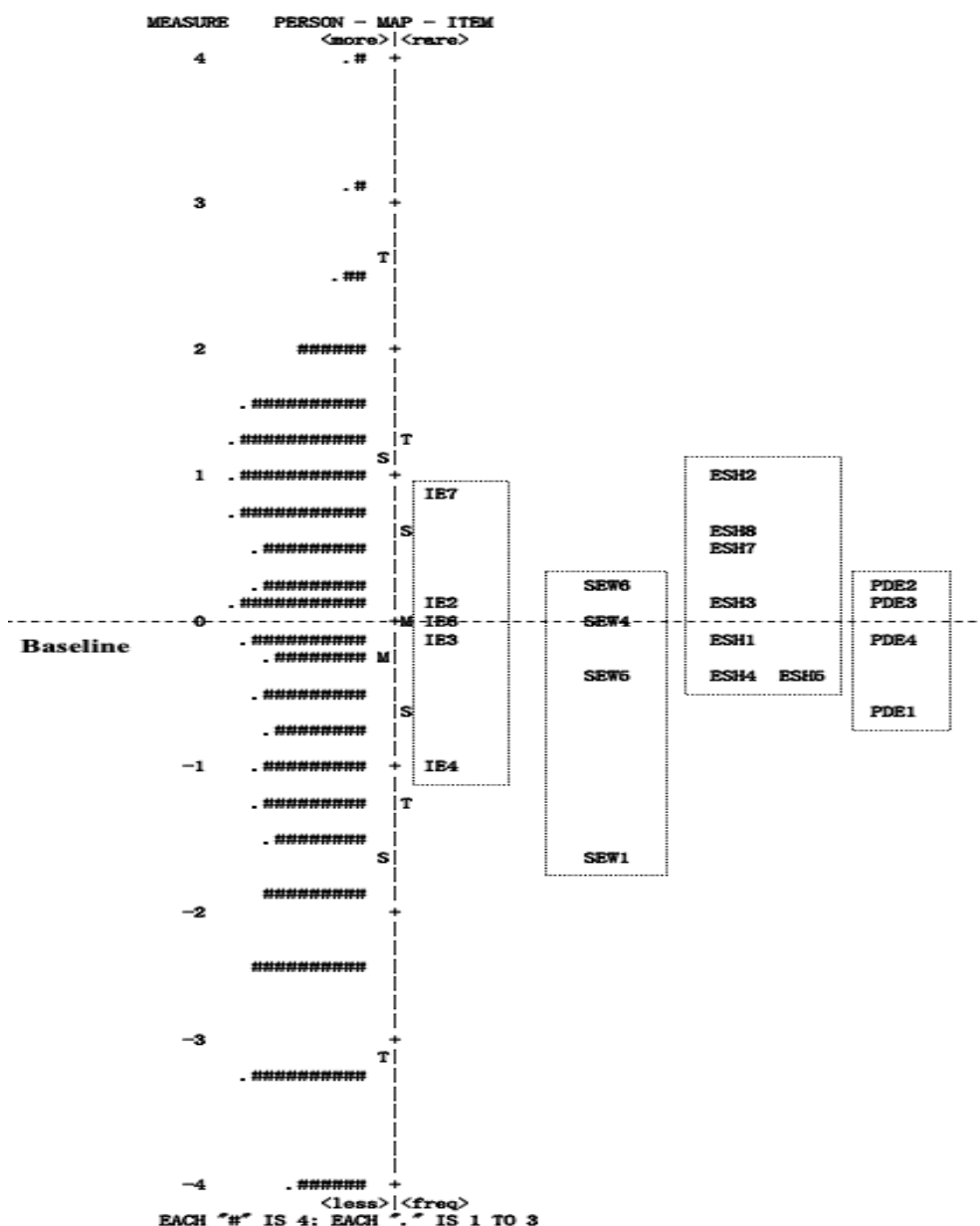


Table 4 shows that the mean of person estimates (-0.23 logit) was lower than the mean of item estimates (0 logits), indicating that the items were difficult for students in the sample. According to Figure 4 and Table 4, the

item difficulties in each aspect were evenly around the baseline (0 logits), which meets the instrument's diagnostic purpose. In particular, the mean of item estimates (0.20) for ESH was higher than the other three aspects, indicating the large number of ACs might exist in the ESH concept. Whereas SEW items had a lower mean of item estimates (-0.46), students might easily understand those items. Another information was the baseline in the person-item map, which could identify the potential items that were more likely to produce students' ACs. The Rasch model assumes that a person's ability at the baseline has a 50% probability of correctly answering the item with its estimate. Persons with lower baseline abilities have lower chances to answer items with higher baseline estimates. Table 4 shows that students' ACs were more likely to be found in the items with estimates above baseline in each concept aspect. Overall, the range of item difficulties was consistent with students' diverse abilities. Therefore, the instrument was validated as a reliable and valid diagnostic tool to measure students' understanding of the CEAS.

Table 4*Mean and Distribution of Item Estimates*

| | Mean of item estimates | Range of item estimate | Items (Above baseline) | Items (Below baseline) |
|-----|------------------------|------------------------|---------------------------|---------------------------|
| IE | -0.02 | [-1.00, 0.95] | IE2, 7 | IE3, 4, 6 |
| SEW | -0.46 | [-1.65, 0.24] | SEW6 | SEW1,4,5 |
| ESH | 0.20 | [-0.40, 1.07] | ESH2, 3, 7, 8 | ESH1, 4, 5 |
| PDE | -0.10 | [-0.66, 0.29] | PDE2, 3 | PDE1, 4 |

Students' Alternative Conceptions about the CEAS

Table 5 presents the percentages of Grade 11 students' correct responses on concept tier, reason tier, and combined tiers, respectively. Table 5 shows that about 50% of students could answer the first tier in each item correctly. In contrast, the correct percentage for the second tier varied by different items. The results suggest that students performed better on the first tier than the second tier in most CEAS-T items. The result also indicates that students might not fully understand the concepts since they could not provide correct reasons for their choices. Besides, items with low percentages of the correct combined tiers need to be examined to explore students' ACs further.

Table 5*The Correct Percentages for First, Second, and Combined Tiers*

| Item | First-tier correct percentage (%) | Second-tier correct percentage (%) | Combined tiers correct percentage (%) |
|------|-----------------------------------|------------------------------------|---------------------------------------|
| IE2 | 54.10 | 60.80 | 42.30 |
| IE3 | 54.00 | 52.40 | 48.40 |
| IE4 | 69.60 | 77.30 | 62.00 |
| IE6 | 69.30 | 56.10 | 46.40 |
| IE7 | 42.00 | 34.40 | 29.60 |
| SEW1 | 78.40 | 75.30 | 71.50 |
| SEW4 | 65.90 | 52.90 | 46.00 |
| SEW5 | 57.20 | 58.00 | 52.00 |
| SEW6 | 71.90 | 48.40 | 41.20 |
| ESH1 | 57.60 | 53.70 | 47.30 |
| ESH2 | 57.20 | 44.90 | 27.70 |
| ESH3 | 59.90 | 51.10 | 43.50 |
| ESH4 | 61.20 | 61.70 | 51.20 |
| ESH5 | 59.30 | 58.00 | 51.90 |
| ESH7 | 42.70 | 46.00 | 35.90 |
| ESH8 | 45.30 | 52.40 | 35.30 |

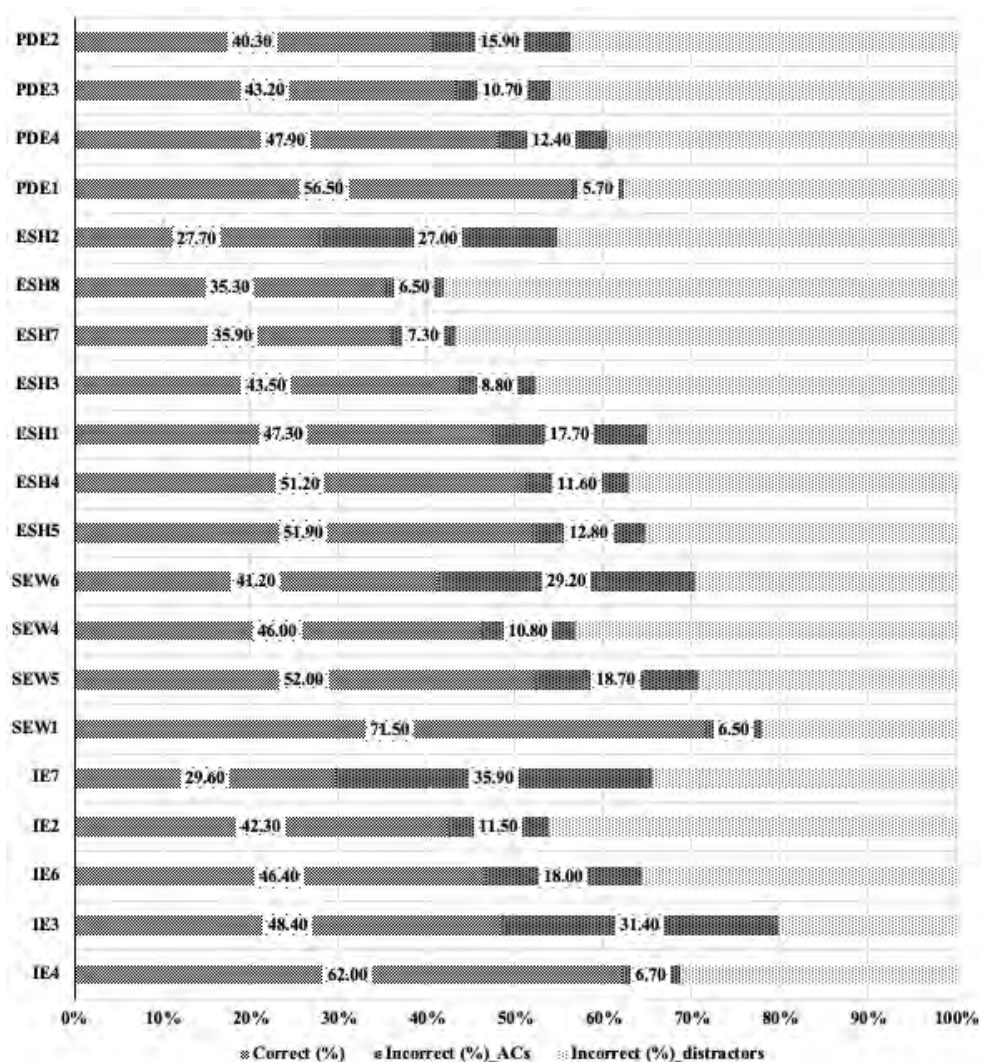


| Item | First-tier correct percentage (%) | Second-tier correct percentage (%) | Combined tiers correct percentage (%) |
|------|-----------------------------------|------------------------------------|---------------------------------------|
| PDE1 | 62.80 | 70.00 | 56.50 |
| PDE2 | 51.70 | 50.10 | 40.30 |
| PDE3 | 51.10 | 51.20 | 43.20 |
| PDE4 | 59.70 | 53.90 | 47.90 |

Students' ACs were identified through students' incorrect responses. Figure 5 presents the percentages of correct and incorrect answers for each item. The items were ordered by the correct rates from high to low for each concept. Meanwhile, the incorrect percentage for each item was presented in the upper area of the bars. Figure 5 shows that students might have had varying degrees of ACs in each item. For example, some items (e.g., IE4 and SEW1) had high percentages of correct responses, but quite a few students still answered incorrectly (38% and 28.50%, respectively). Whereas many students might have had ACs in some items such as ESH2 because the correctness for ESH2 was only 27.70%.

Figure 5

The Percentage of Correct and Incorrect Responses for Each Item



According to Peterson (1986), ACs were further detected when the percentage of students' incorrect responses (the combination of concept and reason) was higher than 10%. The perception of ACs in each item can be seen from the middle bars in Figure 5. Students might have had more than one AC in one item, such as item IE3 (AC2 and AC3 in Table 6). Thus, the perceptions of the two ACs (31.40%) were combined in the IE3 item. Six items (items IE4, SEW1, ESH3, ESH7, ESH8, and PDE1) were not included and presented in Table 6 because those items had less than 10% incorrect responses. Thus, 15 ACs related to the CEAS were identified and grouped based on 750 Grade 11 students' responses (see Table 6).

Table 6
Summary of Students' Alternative Conceptions

| Alternative Conceptions | Choice Combination | Percentage |
|--|--------------------------|------------|
| Ionization equilibrium (IE) | | |
| AC1. The conductivity of electrolyte solutions depends on the mass of electrolytes rather than the concentration of electrolytes or the number of ions. | IE2-A (3), D (2) | 11.5 |
| AC2. The ionic equilibrium exists when some insoluble salts (e.g., CaCO_3) dissolve in water because they are weak electrolytes. | IE3-A (2) | 12.1 |
| AC3. No ionic equilibrium exists when some insoluble bases (e.g., $\text{Cu}(\text{OH})_2$) dissolve in water because they are strong electrolytes. | IE3-D (3) | 19.3 |
| AC4. Adding water into a weak acidic solution promotes the shifting of ionic equilibrium, thus increasing the concentration of hydrogen ions. | IE6-A (2), C (4) | 18.0 |
| AC5. When strong and weak acids with the same pH react with the same metal, the former reaction produces more H_2 gas than the latter. | IE7-C (1) | 35.9 |
| Self-ionization equilibrium of water (SEW) | | |
| AC6. The pH value of water increases because the value of K_w increases when the water is heated. | SEW4-C (2) | 10.8 |
| AC7. At room temperature (25 °C), the solute in a pH < 7 solution is an acid rather than a salt. | SEW5-A (1) | 18.7 |
| AC8. At room temperature (25 °C), if the water produces 1×10^{-13} mol/L hydrogen cations, the solution is a base rather than a salt or an acid. | SEW6-D (3), B (3) | 29.2 |
| Equilibrium of salt hydrolysis (ESH) | | |
| AC9. When an alkali metal salt (e.g., Na_2CO_3) solution and its corresponding base (e.g., NaOH) solution have the same pH value, they have the same degree of equilibrium. | ESH1-A (4) | 17.7 |
| AC10. The degree of hydrolysis of salt solutions depends on the concentrations of hydrogen ions or hydroxide ions in the solution. | ESH2-B (3), D (3) | 27.0 |
| AC11. When a salt (e.g., NaHCO_3) dissolve in water, the final solution does not have any ions (e.g., OH) that are formed by hydrolysis. | ESH4-A (4), C (2), C (4) | 11.6 |
| AC12. When a chemical reaction is done completely in an aqueous solution (e.g., CO_2 reacts with NaOH), the final solution does not have any ions (e.g., OH) that are formed by hydrolysis. | ESH5-C (3) | 12.8 |
| Precipitation and dissolution equilibrium (PDE) | | |
| AC13. Adding more solid solutes (e.g., BaCrO_4) to its' solution, the precipitation and dissolution equilibrium changes. | PDE2-B (2) | 15.9 |
| AC14. The precipitation and dissolution equilibrium only exist in a poorly soluble electrolyte (e.g., CaCO_3) solution but not in a soluble electrolyte solution (e.g., saturated NaCl). | PDE3-D (2) | 10.7 |
| AC15. The solubility product constant (K_{sp}) changes at a specific temperature when the precipitation and dissolution equilibrium changes. | PDE4-B (4) | 12.4 |



Discussion

The primary goal of this study was to develop a valid and reliable two-tier multiple-choice diagnostic instrument (CEAS-T; see the final test in Appendix C) to assess students' understanding of the CEAS. Using Rasch analysis, this study presented psychometric evidence to support the validity and reliability of the CEAS-T instrument for measuring students' ACs about the CEAS. The CEAS-T instrument was further employed to investigate ACs about the CEAS of 750 Chinese Grade 11 students. In total, 15 ACs about the CEAS were identified when the percentage of incorrect combinations of the two tiers was higher than 10% (Peterson 1986).

Validity, Unidimensionality, and Reliability

This study employed Treagust's (1988) framework for designing two-tier multiple-choice diagnostic items to ensure the CEAS-T instrument's validity. Consistent to previous development studies (Cheong et al. 2015; Karpudewan et al. 2015), this study supports the robust design principle (Treagust, 1988) to design a new diagnostic instrument that is aligned with the content standards in Chinese curriculum standards. Moreover, the Treagust's (1988) framework and external review process were used to support the content and face validity of the CEAS-T instrument (Trochim & Donnelly 2006). Instead of using the CTT-based validation method, this study employed an IRT-based validation process to provide further psychometric evidence for validating the developed instrument. Our study reinforces the benefit of using IRT-based method to validate two-tier instruments (Fulmer et al., 2015; Lu & Bi 2016; Romine et al., 2015). As an IRT-based approach, Rasch modeling was utilized to provide evidence for further validating the CEAS-T instrument. The unidimensionality of the instrument was supported by considerable evidence, such as the PCA, eigenvalue of the unexplained variance, and items' local independence. The item fit indices and item-person mapping findings provided evidence at the item and overall levels to show the validity of the instrument. Three of those items (PDE1, 3, and 4) were probably designed for other constructs, which should be further considered. Those items were about the topic of precipitation and dissolution equilibrium. One possible reason for this could be that the precipitation and dissolution equilibrium is about saturated solutions, whereas the other equilibriums in aqueous solutions are about unsaturated solutions.

Students' Alternative Conceptions about the CEAS

This study builds upon previous AC research about chemical equilibrium (Akkus et al., 2011; Karpudewan et al., 2015) and solution chemistry (Damanhuri et al., 2016; Lu & Bi, 2016; Orwat et al., 2017) to explore student ACs in a particular case of chemical equilibrium in aqueous solutions (CEAS). Using the developed instrument, this study found that most students had better performance on concept tiers than reasoning tiers. These findings are consistent with previous studies (e.g., Adadan & Savasci, 2012; Artdej et al., 2010) that students could easily answer the content of a specific concept but lacked the understanding of the nature of those concepts (Adadan & Savasci, 2012). This study also found that six items (item IE4, SEW1, ESH3, ESH7, ESH8, and PDE1) could not detect students' ACs. One reason could be that most students had a good performance on those items, supported by the high correct percentages (e.g., 71.50% for SEW1 in Figure 5). Another possible reason could be the design of distractor options was not sensitive enough to detect students' ACs. Although they had excellent fit indices, the high incorrect percentages for unknown combinations were still found (e.g., 47.70% for ESH3 in Figure 5). Therefore, those items need to be further re-examined and revised.

Some findings are similar with previous research, especially in the following subtopics, such as distinguishing strong and weak electrolytes (Artdej et al., 2010; Damanhuri et al., 2016); calculating pH value for a solution by using Kw (Artdej et al., 2010; Demircioğlu et al., 2005); identifying all species in a solution with the equilibrium of salt hydrolysis (Orwat et al., 2017); and figuring out the factors affecting the precipitation and dissolution equilibrium (Bilgin & Geban, 2006). However, different from those studies, the findings indicated that students had difficulty making connections among those concepts, which is critical for students' deeper understanding (Clark & Linn, 2013; Davis, 2000; NRC, 1999). One challenge is mathematical thinking towards transforming concentration, the equilibrium constant, and pH value. Another is building the relationships among acidity, solubility, ionization, and chemical reaction. The findings of this study indicated that the CEAS is a complex construct for students to learn, particularly in building relationships among multiple related concepts. Therefore, teachers



should pay additional attention to support students' deeper learning by building integrated knowledge among various related CEAS concepts (Linn, 2006).

Limitations

This study does have several limitations. First, the convenience sampling method is not robust enough to represent student population in mainland China. Future studies will be conducted using more rigorous sampling strategies such as systematic sampling or cluster sampling strategies. Second, some items may not be sensitive enough to diagnose students' ACs for the topic. As an iterative process of Rasch analysis (He et al., 2016; Liu & Boone, 2006), the revision and validation of the CEAS-T instrument should be further conducted. Third, both qualitative and quantitative data are needed to explain those ACs about the CEAS. Although several qualitative analyses during the design process were conducted, this study did not conduct follow-up interviews or observations. Without rich, qualitative data sources (McClary & Bretz, 2012), it is impossible to interpret students' conceptions in the CEAS.

Implications

The broad implication of this study is leveraging the findings of those ACs to inform innovative instructional strategies, especially in a topic-specific manner. For example, when teaching the topic of chemical equilibrium in aqueous solutions, teachers should take advantage of students' prior knowledge of chemical equilibrium and solution chemistry to build up a strong integrated knowledge (Linn, 2006). Moreover, integrating those concepts should be taught explicitly with meaningful context (McClary & Bretz, 2012). For understanding the topic quantitatively, classroom activities are desired to incorporate mathematical representations (Waldrup & Prain, 2012). For instance, graphs alongside equations can help students obtain calculation ability in figuring out the transformation of concentration, pH value, and equilibrium constant (Park & Choi, 2013). Virtual learning environments (Linn, 2006) that enhance students' learning at both macroscope and microscope levels would be preferred for students making sense of the CEAS. Moreover, students would benefit from technology-enhanced activities (Linn, 2006) to better understand acid-base titration, such as handheld-based science experiments (Roschelle et al., 2005). As such, more studies are needed to explore whether those innovative instructional strategies can improve students' understanding of such complex constructs (e.g., the CEAS).

The present study also raises the possibility of applying the CEAS-T instrument as a reliable and valid diagnostic assessment tool for formative and summative purposes (Maier et al., 2016; Mutlu & Sesen, 2016). For formative purposes, the items in the developed instrument can be selected purposely to match the targeted lessons or units. For example, students' ACs on previous studies could be used as prior knowledge for teachers to adjust the following lesson plans. For summative purposes, the CEAS-T instrument can assess students' learning achievement after studying as completing their learning units related to the topic.

Conclusions

In conclusion, this study developed a valid and reliable two-tier multiple-choice instrument to measure students' ACs toward the CEAS and employed this instrument to detect a handful of common ACs in a sample of Chinese upper secondary students. Considering the complexity of the CEAS, we believe our instrument would help worldwide researchers who are interested in diagnosing upper secondary students' ACs toward the related topics of the CEAS. We also believe the detected common ACs could be as useful resources for teachers to plan and adjust their instructional design to better support students' learning on such complex concepts.

Acknowledgements

This research is based upon work supported by several grants, such as the National Education Sciences Planning Program (General Category) under Grant No. BHA170131, Grant No. NENU-421/131005031 and No. NENU-17XQ008. Any opinions, findings, and conclusions or recommendations expressed in the materials are those of the authors and do not necessarily reflect the views of those funding grants.



Declaration of Interest

The authors declare no competing interest.

References

- Adadan, E., & Savasci, F. (2012). An analysis of 16–17-year-old students' understanding of solution chemistry concepts using a two-tier diagnostic instrument. *International Journal of Science Education*, 34(4), 513-544. <https://doi.org/10.1080/09500693.2011.636084>
- Akkus, H., Kadayifci, H., & Atasoy, B. (2011). Development and application of a two-tier diagnostic test to assess secondary students' understanding of chemical equilibrium concepts. *Journal of Baltic Science Education*, 10(3), 146-155. <http://oaji.net/articles/2014/987-1410547169.pdf>
- Artdej, R., Ratanaroutai, T., Coll, R. K., & Thongpanchang, T. (2010). Thai grade 11 students' alternative conceptions for acid–base chemistry. *Research in Science & Technological Education*, 28(2), 167-183. <https://doi.org/10.1080/02635141003748382>
- Bilgin, I., & Geban, Ö. (2006). The effect of cooperative learning approach based on conceptual change condition on students' understanding of chemical equilibrium concepts. *Journal of Science Education and Technology*, 15(1), 31. <https://doi.org/10.1007/s10956-006-0354-z>
- Bond T. G., & Fox C. M., (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.), Lawrence Erlbaum.
- Çalik, M. (2005). A cross-age study of different perspectives in solution chemistry from junior to senior high school. *International Journal of Science and Mathematics Education*, 3(4), 671-696. <https://doi.org/10.1007/s10763-005-1591-y>
- Cheong, I. P. A., Johari, M., Said, H., & Treagust, D. F. (2015). What do you know about alternative energy? Development and use of a diagnostic instrument for upper secondary school science. *International Journal of Science Education*, 37(2), 210-236. <https://doi.org/10.1080/09500693.2014.976295>
- Clark, D. B., & Linn, M. C. (2013). The Knowledge integration perspective: Connections across research and education. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (2nd ed., pp. 61-82). <https://doi.org/10.4324/9780203154472.ch27>
- Damanhuri, M. I. M., Treagust, D. F., Won, M., & Chandrasegaran, A. L. (2016). High school students' understanding of acid-base concepts: An ongoing challenge for teachers. *International Journal of Environmental and Science Education*, 11(1), 9-27. <https://doi.org/10.12973/ijese.2015.284a>
- Davis, E. A. (2000). Scaffolding Students' Knowledge Integration: Prompts for Reflection in KIE. *International Journal of Science Education*, 22(8), 819-837. <https://doi.org/10.1080/095006900412293>
- Demircioğlu, G., Ayas, A., & Demircioğlu, H. (2005). Conceptual change achieved through a new teaching program on acids and bases. *Chemistry Education Research and Practice*, 6(1), 36-51. <https://doi.org/10.1039/B4RP90003K>
- Fulmer, G. W., Chu, H. E., Treagust, D. F., & Neumann, K. (2015). Is it harder to know or to reason? Analyzing two-tier science assessment items using the Rasch measurement model. *Asia-Pacific Science Education*, 1(1), 1. <https://doi.org/10.1186/s41029-015-0005-x>
- Garnett, P. J., Garnett, P. J., & Hackling, M. W. (1995). Students' Alternative Conceptions in Chemistry: A Review of Research and Implications for Teaching and Learning. *Studies in Science Education*, 25 (1), 69-95. <https://doi.org/10.1080/03057269508560050>
- He, P., Liu, X., Zheng, C., & Jia, M. (2016). Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. *Chemistry Education Research and Practice*, 17(2), 381-393. <https://doi.org/10.1039/C6RP00004E>
- Karpudewan, M., Treagust, D. F., Mocerino, M., Won, M., & Chandrasegaran, A. L. (2015). Investigating high school students' understanding of chemical equilibrium concepts. *International Journal of Environmental and Science Education*, 10(6), 845-863. <https://doi.org/10.12973/ijese.2015.280a>
- Kousathana, M., Demerouti, M., & Tsaparlis, G. (2005). Instructional misconceptions in acid-base equilibria: An analysis from a history and philosophy of science perspective. *Science & Education*, 14(2), 173-193. <https://doi.org/10.1007/s11191-005-5719-9>
- Licht, P. (1991). Using a Diagnostic Test of Pupils' Alternative Conceptions to Plan a Teaching Strategy on Electric Circuits. *European Journal of Teacher Education*, 14(1), 19-30. <https://doi.org/10.1080/0261976910140104>
- Linn M. C. (2006). The knowledge integration perspective on learning and instruction. In R. K. Sawyer (ed.), *The Cambridge handbook of the learning sciences* (pp. 243–264). Cambridge University Press.
- Linacre, J. M., & Tennant, A. (2009). More about critical eigenvalue sizes in standardized-residual principal components analysis (PCA). *Rasch Measurement Transactions*, 23(3), 1228. <https://www.rasch.org/rmt/rmt233f.htm>
- Liu, X., & Boone, W. J. (2006). *Applications of Rasch measurement in science education*. JAM Press.
- Lu, S., & Bi, H. (2016). Development of a measurement instrument to assess students' electrolyte conceptual understanding. *Chemistry Education Research and Practice*, 17(4), 1030-1040. <https://doi.org/10.1039/C6RP00137H>
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, 95, 85-98. <https://doi.org/10.1016/j.compedu.2015.12.002>



- McClary, L. M., & Bretz, S. L. (2012). Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *International Journal of Science Education*, 34(15), 2317-2341. <https://doi.org/10.1080/09500693.2012.684433>
- McCoach D. B., Gable R. K., & Madura J. P. (2013). Instrument development in the affective domain. Springer. <https://doi.org/10.1007/978-1-4614-7135-6>
- Mdachi, S. J. (2012). Giving Thought to Students' Alternative Conceptions in Stereochemistry: One Teacher's Basis for Pedagogical Content Knowledge Improvement. *Journal of Turkish Science Education*, 9(4), 22-34.
- Ministry of Education, P. R. China. (2017). 普通高中化学课程标准 [Chemistry curriculum standards for senior high school]. People's Education Press.
- Morales, R. A. (2009). Evaluation of mathematics achievement test: A comparison between CTT and IRT. *The International Journal of Educational and Psychological Assessment*, 1(1), 19-26.
- Mutlu, A., & Şeşen, B. A. (2016). Evaluating of preservice science teachers' understanding of general chemistry concepts by using two tier diagnostic test. *Journal of Baltic Science Education*, 15(1), 79. <https://dx.doi.org/10.33225/jbse/16.15.79>
- National Research Council. (1999). *How people learn: Bridging research and practice*. The National Academies Press. <https://doi.org/10.17226/9457>
- Odden, T. O. B., & Russ, R. S. (2019). Defining sensemaking: Bringing clarity to a fragmented theoretical construct. *Science Education*, 103(1), 187-205. <https://doi.org/10.1002/sce.21452>
- Orwat, K., Bernard, P., & Migdał-Mikuli, A. (2017). Alternative conceptions of common salt hydrolysis among upper-secondary-school students. *Journal of Baltic Science Education*, 16(1), 64-76. <https://dx.doi.org/10.33225/jbse/17.16.64>
- Özmen, H. (2008). Determination of students' alternative conceptions about chemical equilibrium: A review of research and the case of Turkey. *Chemistry Education Research and Practice*, 9(3), 225-233. <https://doi.org/10.1039/B812411F>
- Padilla J. L., and Leighton J. P., (2017). Cognitive interviewing and think aloud methods. In B. D. Zumbo & A. M. Hubley (Eds.), *Understanding and investigating response processes in validation research* (1st ed., pp. 211-228). Springer. <https://doi.org/10.1007/978-3-319-56129-5>
- Park, E. J., & Choi, K. (2013). Analysis of student understanding of science concepts including mathematical representations: pH values and the relative differences of pH values. *International Journal of Science and Mathematics Education*, 11(3), 683-706. <https://doi.org/10.1007/s10763-012-9359-7>
- Peterson R.F. (1986). The development, validation, and application of a diagnostic test measuring year 11 and 12 students' understanding of covalent bonding and structure, [Unpublished Master's thesis]. Curtin University of Technology.
- Piquette, J. S., & Heikkinen, H. W. (2005). Strategies reported used by instructors to address students' alternate conceptions in chemical equilibrium. *Journal of Research in Science Teaching*, 42(10), 1112-1134. <https://doi.org/10.1002/tea.20091>
- Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Measurement Transactions*, 19(1), 1012. <https://www.rasch.org/rmt/rmt191h.htm>
- Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment*, 84(3), 228-238. https://doi.org/10.1207/s15327752jpa8403_02
- Romine, W. L., Schaffer, D. L., & Barrow, L. (2015). Development and application of a novel Rasch-based methodology for evaluating multi-tiered assessment instruments: Validation and utilization of an undergraduate diagnostic test of the water cycle. *International Journal of Science Education*, 37(16), 2740-2768. <https://doi.org/10.1080/09500693.2015.1105398>
- Roschelle, J., Penuel, W. R., Yarnall, L., Shechtman, N., & Tatar, D. (2005). Handheld tools that 'informate' assessment of student learning in science: a requirements analysis. *Journal of Computer Assisted Learning*, 21(3), 190-203. <https://doi.org/10.1109/WMT.2004.1281367>
- Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, 28(9), 761-784. <https://doi.org/10.1002/tea.3660280905>
- 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. <https://doi.org/10.1002/tea.10023>
- 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. <https://doi.org/10.1080/0950069880100204>
- Trochim W. M., & Donnelly J. P. (2006). *The research methods knowledge base* (3rd ed.). Atomic Dog.
- Voska, K. W., & Heikkinen, H. W. (2000). Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, 37(2), 160-176.
- Waldrup B., & Prain V. (2012). Learning from and through representations in science. In B. J. Fraser, K. Tobin, & C. J. McRobbie (Eds), *Second international handbook of science education* (1st ed., pp. 145-155). Springer. <https://doi.org/10.1007/978-1-4020-9041-7>
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Yang, Y., He, P., & Liu, X. (2018). Validation of an instrument for measuring students' understanding of interdisciplinary science in grades 4-8 over multiple semesters: A Rasch measurement study. *International Journal of Science and Mathematics Education*, 16(4), 639-654. <https://doi.org/10.1007/s10763-017-9805-7>



Appendix

Appendices A, B, C can be found in a repository via the URL below: <https://drive.google.com/file/d/1Tkcr7jIkAmzgTmDWg7WCRyT68vWnrqRX/view?usp=sharing>

Received: December 18, 2021

Revised: March 18, 2022

Accepted: May 12, 2022

Cite as: He, P., Zheng, C., & Li, T. (2022). Upper secondary school students' conceptions of chemical equilibrium in aqueous solutions: Development and validation of a two-tier diagnostic instrument. *Journal of Baltic Science Education*, 21(3), 428-444. <https://doi.org/10.33225/jbse/22.21.428>

Peng He
(Corresponding author)

PhD, Assistant Professor, Institute of Chemistry Education, College of Chemistry, Northeast Normal University, China; Research Associate, CREATE for STEM Institute, College of Education, Michigan State University, USA.

E-mail: hep905@nenu.edu.cn

ORCID: <https://orcid.org/0000-0002-2877-0117>

Changlong Zheng
(Corresponding author)

Professor, Institute of Chemistry Education, College of Chemistry, Northeast Normal University, China.

E-mail: zhengcl@nenu.edu.cn

ORCID: <https://orcid.org/0000-0003-2229-9843>

Tingting Li

PhD, Faculty of Education, Northeast Normal University, Changchun, China; PhD student, Department of Counseling, Psychology, and Special Education, College of Education, Michigan State University, USA

E-mail: litt606@nenu.edu.cn

ORCID: <https://orcid.org/0000-0002-5692-2042>

