Competency model of designing competency assessment tool: A pilot study with Vietnamese science pre-service teacher

Do Huong Tra* and Nguyen Thi Dieu Linh

Hanoi National University of Education (HNUE), 136 Xuan Thuy Road, Cau Giay District, Hanoi, Vietnam
Corresponding authors E-mail: 1 dhtra@hotmail.com; 2 linhndt@hnue.edu.vn

Abstract. Assessment in STEM education is a difficult activity. It requires teachers to have the competence of designing a competency assessment tool (CDCAT). This article proposes the model of CDCAT. The components and behavioral indicators of the CDCAT were proposed based on analyzing the thinking process when designing an assessment tool. The model was built through a rigorous multi-step research process that ensures the value and reliability of the model. The pilot of 53 science pre-service teachers also showed that the model has a guiding effect on the design of the assessment tool in teaching science.

Keywords: assessment tool, competence, tool designing, competency assessment, competence model.

1. Introduction

STEM education is a modern trend in education. To help students develop 21st-century skills in STEM education, teachers should be equipped with knowledge and skills to select, adapt, and design of classroom assessment tool per curriculum-based competence [44], [45], [46]. The meta-analysis of 250 empirical studies conducted [1] showed that using formative assessment in teachers’ everyday teaching practice resulted in substantial gains in student achievement scores on standardized tests and that is larger than most other types of impact on education [1].

Previous studies on classroom assessments have shown that many teachers are inadequately trained and ill-prepared to design and develop an assessment tool [2-4]. Empirical studies in some countries have also found that the majority of the teacher candidates reported a low level of assessment competence including CDCAT and expressed the need for improving their assessment competence through specific courses in classroom assessment and testing [5] [6].

In Vietnam, although there have been training by the Vietnamese Ministry of Education and Training on the innovation of assessment and testing competencies for teachers, they still have many difficulties in designing a competency assessment tool [7]. This shows an important role in initial training at pedagogical universities.

Recognizing the importance of innovation in assessment and testing in general, and CDCAT in particular, some pedagogical universities have included the module in their curricula. However, the training and developing of any competencies should be based on the model of that competence [8]. Therefore, developing the CDCAT model is very necessary.

There are many assessment literacy standards for teachers in the world and there are also many studies on the models of competencies assessment [9-20]; however, the CDCAT is just a component in such models. According to our investigation results, no research has focused on clarifying the CDCAT...
model. To help lecturers determine objectives of the activity to develop for pre-service teachers’ CDCAT, this research recommends the CDCAT model built up from the given output results.

2. Research Methods
Based upon a review of literature on assessment design and competence model construction the research has used a multi-step development process to develop the CDCAT model that meets the content and structural validity [8] [42]. In particular, we have:

- analyzed the literature on the model of assessment competencies, assessment procedures, assessment literacy standards for teachers in the past and at present to determine components and behavioral indicators of the model,
- consulted experts (for the first time) on the components and behavioral indicators of the model and conducted survey on teachers or pre-service teachers to determine the content validity of the initial model,
- reviewed the literature on the evaluation criteria, the levels assessed of assessment tools and practical data collected during the training and retraining the CDCAT for in and pre-service teachers, in combination with interviews on pre-service teachers’ cognition to propose quality criteria, consulted experts (for the second time) on the complete competence model of designing competency assessment tools,
- designed assessment tasks to evaluate the CDCAT; piloting assessment tasks and conducting a trial on science pre-service teachers.

3. Content and research results
3.1. Hypotheses about competency model of designing competency assessment tool

3.1.1. The concept of competence of designing competency assessment tool. Assessment competence is related to the understanding and appropriate use of assessment practices, theoretical and philosophical background in measuring student learning [21][6]. Another simpler definition is provided by the North Central Regional Educational Laboratory: “Assessment competence is the willingness of an educator to design, implement and discuss assessment strategies” [22]. From the above definition, it can be seen that the CDCAT is a component of assessment competence.

To define and clarify the CDCAT model, we analyzed the Stiggins assessment competence model, and 12 assessment literacy standards for teachers (from 1990 to present) from six geographic regions (USA, Canada, UK, Europe, Australia and New Zealand) [9-20]. Through this analysis, we have defined “competence of designing competency assessment tool as the capacity to select appropriate assessment methods and develop assessment exercises and appropriate scoring procedures for assessment purposes”.

3.1.2. Proposal of the competence model of designing competency assessment tool. To propose the hypothesis about the competency model, it is necessary to identify the construct and describe the construct in terms of observable evidence [8]. Stiggins’ model of assessment competence and 12 assessment literacy standards for teachers demonstrate the sequence of steps to be taken during designing a competency assessment tool [9-20]. Thereby, it is possible to determine the components of the CDCAT; however, the behavioral indicators have not been identified.

This study analyses the processes of designing assessments, which are popularly used in research [23-30]. To draw out behavioral indicators for the CDCAT model, we compare actions for each step of one process to those for another. Through this analysis, we find out the common and reasonable points of processes, then analyze and add some points as follows:

- Each assessment activity should be contextualized and diversified in purposes. To ensure that, it is necessary to specify the purposes into specific goals. Through that, the design of an assessment tool is better oriented. Among the processes analyzed, only the process of Tillema (2011) pays attention to specifying the goals but does not indicate the aspects to be considered when determining goals. Therefore, at this stage, after determining the purpose of the assessment, we include two steps: (1) Identifying the characteristics of the situations; (2) Determining the objectives of the assessment tasks.
It is difficult to select situations in a competency assessment tool that must meet the key criteria on Accuracy, Generality, and Extrapolation [31]. The above processes have not paid attention to this complexity. When analyzing the thinking process in designing exercises, we have added the following two steps before the step of designing the assessment task: (1) Specifying the type of information to be used; (2) Searching for the type of information to be used.

Corresponding to the steps in the assessment designing process, behavioral indicators in the CDCAT model were proposed as shown in Figure 1.

![Figure 1. Initial hypothetical model of the CDCAT](image)

Because the selected literature does not yet clarify quality criteria, we propose a tentative model, just including components and behavioral indicators.

Determining the quality criteria will not make sense when the components and behavioral indicators are not appropriate. Therefore, this model has been sent to five colleagues for feedback. They are active lecturers who have taught assessment and testing modules or have had research in assessment.

Three experts agreed with all the proposed behavioral indicators. Only two experts suggested separating the behavior indicator 12 into the following behavioral indicators: coding and analyzing. However, it is possible to see that coding is defined in the behavioral indicator 10. Before implementing the behavior indicator 12, coding is just a set of predefined steps that do not demonstrate the use of assessment knowledge. Only analyzing activities should use knowledge of reliability, validity, difficulty, discrimination and some data processing software. Therefore, we did not split the behavioral indicators and still retained the model as shown in Figure 1.

3.2. Determining the structural validity of the hypothetical model of the CDCAT

To obtain evidence for the reliability of the CDCAT model in Figure 1, a survey has been conducted on 60 science pre-service teachers at HNUE. Each student was asked to assess the level of confidence for each behavioral indicator of the CDCAT. Before the investigation, the questionnaire was tested on 15 pre-service teachers to correct points that could be misunderstood.

To determine the structural value, the Cronbach's Alpha coefficient was initially calculated for all 13 behavioral indicators (Figure 1). However, the variable VAR0007 corresponding to the behavior indicator 7 had small Corrected item-total Correlation. When removing this variable, Cronbach's Alpha increased.

It can be seen that the behavioral indicator "7. Search for the type of information to be used" is necessary for the process of designing a competency assessment tool. However, a person who can
search for information but does not have the necessary knowledge about assessment cannot design a competency assessment tool. A person with an average ability to search for information can do this if he/she can use the knowledge of assessment. Thus, the experimental results are perfectly appropriate. We have chosen a process-based approach to determine the behavioral indicators of the competency model of designing competency assessment. From this analysis, it can be seen that keeping track of the cognitive processes, trying to introduce the full stages of the designing process without appropriately considering the necessary knowledge is a mistake. Right after that, with each behavioral indicator, we proceeded to identify all the necessary knowledge to carry out those behaviors to avoid the above mistake. As a result, the remaining behavioral indicators did not have the same problems as the indicator 7.

After removing the VAR0007 variable, the analysis results show that the Cronbach’s Alpha coefficient is 0.877, in the range from 0.8 to 1, which means that the scale is very good. The Corrected item-total Correlation coefficient of all behavioral indicators is greater than 0.3 (from 0.868 to 0.876), which means all the behavioral indicators are satisfactory; no other behavioral indicators measure another competence.

The statistical analysis shows that the above 12 behavioral indicators are well correlated with each other and measure the same variable, the CDCAT of pre-service teachers.

3.3. Determine the quality criteria of the hypothetical model of CDCAT

The first draft of the quality criteria needs to built based on the combined knowledge gained from reviewing the literature on learning outcomes, and recommended evaluation criteria for the assessment methods concerned; analysis of student artifacts; and in-depth discussions with instructors [42]. Therefore, to propose quality criteria, we reviewed the evaluation criteria and the levels assessed in 9 assessment tools [32-40], in combination with the practical data below.

In the summer of 2014, the Ministry of Education of Vietnam organized training courses on the assessment of teachers’ competencies in the whole country. Each province has 5 to 7 teachers. A survey of 382 teachers was conducted to identify the level of difficulty that teachers encountered designing competency assessment tools. The assessment tools designed by those teachers, and our observations during the teacher training process and the results of the survey are sources of practical data from teachers as a basis for proposing quality criteria. [7].

In the academic year 2015-2016, in the module “Assessment and Testing”, 56 pre-service teachers of Physics Faculty, HNUE, were required to make their portfolio. Before each class, pre-service teachers had to read the documents of the lesson and determine the objectives of that lesson. After class, pre-service teachers recorded what they had learned, the difficulties they encountered, and what to learn more in their portfolio. The assessment tool designed by pre-service teachers, observations in the teaching process and the academic records are practical data from pre-service teachers as a basis for proposing quality criteria.

In the three consecutive years (2015-2017), to obtain deeper insight into pre-service teachers’ perceptions when designing assessment tool, we collected data during the process of supervising 9 physics pre-service teachers to do graduation theses on the topic of designing competency assessment tool. We discussed directly with pre-service teachers to find out their difficulties, suggest directions and monitor changes during their process of solving problems; directly interviewing the pre-service teachers’ cognitive processes every time they designed new tools. The tools that pre-service teachers designed during their graduation theses and the above data also help us to propose quality criteria.

In-depth discussions were held with two authors of the article and two lecturers who have taught and researched in assessment with the above data. Based on these discussions we chose SOLO taxonomy [41] to propose the quality criteria. We use this rating scale because it is also a simple, reliable and easy-to-use model that matches the quality rating of those behavioral indicators.

To get the feedback on the model, we consulted experts the second time. We select 23 experts who are active lecturers teaching the assessment and testing module at universities of education and experts on assessment from quality assurance and accreditation centers to collect feedback on the quality criteria. The experts were asked to rate and comment on the clarity,
completeness and generalized applicability of the quality criteria. With the points those experts disagree, both in terms of content and expression, we suggest that they provide a measure to adjust.

The two behavioral indicators 6 and 8 (Figure 1) require creativity. According to the rubric of Alonzo, corresponding to each behavior, we designed the highest level “Instructing the behavior in a professional way for colleagues” [40]. To reach this level, pre-service teachers need to draw up clearly, logical ways and rules. That also corresponds to the highest level in the SOLO taxonomy. However, 52% of experts said that this was not reasonable because, to achieve those quality criteria, pre-service teachers also would be able to present the information. Other feedbacks from experts focused on expressing to clarify the meanings. 30% of experts thought that indicator 9 should be made clearer in meaning. Some individual responses focused on terms and expressions. Analyzing feedback from experts, we have made some modifications to obtain the quality criteria for the CDCAT as shown in Table 1.

The full model was sent to 4 experts to assess its content validity. The instructors were then asked to reflect upon the following set of questions suggested by Moskal and Leydens (2000): “Do the evaluation criteria of the scoring rubric address all aspects intended to be measured in the assignments given by you? (Content validity); Are all the important criteria relevant to the assessment method being evaluated through the rubric? (Construct validity); and Do the scoring criteria reflect competencies that would suggest success on future or related performances? (Criterion validity)” The answers from the instructors were in the affirmative. [43]

| Components | Behavioral indicators | Quality criteria |
|------------|----------------------|-----------------|
| I. Determine the purposes and objectives of the competency assessment activities | 1. Identify the purposes of using the assessment tool | Md1. State the purpose of using the assessment tool in a general way |
| | | Md2. State clearly, fully some purposes of using the assessment tool |
| | | Md3. State clearly, fully the purposes of using the assessment tool |
| | | Md4. State, classify, and rank the purposes of using the assessment tool |
| | 2. Determine the characteristics of the situation using the tool | Th1. Identify some common factors (time, acquired knowledge) |
| | | Th2. Identify some basic factors, consistent with the assessment purposes |
| | | Th3. Fully determine the factors to be considered, appropriate for assessment purposes |
| | | Th4. Determine, classify and rank the factors to be considered |
| | 3. Determine the objectives of the assessment task system. | Mt1. Determine the objective, but only focusing on a number of common factors |
| | | Mt2. Determine the objective, focusing on the competence model to be assessed |
| | | Mt3. Determine the objective, ensuring the purposes of assessment |
| | | Mt4. Determine the objective that is relevant to the situation in which the tool is being used |
| II. Plan the development of the assessment tool. | 4. Determine the type of evidence to be collected to assess learners’ competencies | Bc1. Determine the type and amount of evidence to assess some behavioral indicators |
| | | Bc2. Determine the type and quantity of evidence to separately assess each behavioral indicator |
| | | Bc3. Determine the type and amount of evidence with attention to the relationship of behavioral indicators |
| | | Bc4. Determine the type and amount of evidence that fully meets the assessment objectives |
| | 5. Select assessment methods supporting the collection of such evidence | Pp1. Select some assessment methods in accordance with the collection of some types of evidence |
| | | Pp2. Select assessment methods supporting the separate collection of each type of evidence |
| | | Pp3. Select assessment methods with attention to the combination of collecting different types of evidence |
| | | Pp4. Select the assessment methods that achieve full assessment objectives |

Table 1: The full model of the CDCAT
Table 1: The full model of the CDCAT

| Components | Behavioral indicators | Quality criteria |
|------------|-----------------------|------------------|
| III. Develop assessment tool | 6. Determine the type of information used to draft assessment tasks | Tt1. Only find assessment tasks that are available for use directly. |
|            |                        | Tt2. Determine the characteristics of the information that can be immediately utilized to draft assessment tasks. |
|            |                        | Tt3. Determine the characteristics of information that can be utilized to find ideas for assessment tasks, alternative to generate other assessment tasks. |
|            |                        | Tt4. Determine the type of information and reflect the implementation process to withdraw appropriate rules for the next time. |
| 7. Draft assessment tasks |  | |
|            |                        | S1. Draft assessment tasks only to obtain some simple assessment objectives. |
|            |                        | S2. Draft assessment tasks to achieve assessment objectives in family situation. |
|            |                        | S3. Draft multi-dimensional assessment tasks that allow students to express their ideas in a variety of ways. |
|            |                        | S4. Draft the assessment tasks and reflect the implementation process to withdraw appropriate rules for the next time. |
| 8. Design assessment instruments to assess the evidence obtained. | | |
|            |                        | Cc1. Select some common kinds of suitable tools |
|            |                        | Cc2. Select the appropriate type of assessment instruments to assess the evidence obtained. |
|            |                        | Cc3. Develop the assessment instruments that meet some basic criteria according to the assessment theory. |
|            |                        | Cc4. Develop appropriate assessment instruments to assess the evidence and create learning opportunities for students. |
| 9. Determine the steps that evaluators should take to use the tool | | |
|            |                        | Sd1. Determine some steps using the tool with reasonable content |
|            |                        | Sd2. Determine steps using the tool with reasonable content |
|            |                        | Sd3. Determine the sequence of steps using the tool is basically reasonable but not optimal |
|            |                        | Sd4. Determine the sequence and content of the steps are optimal |
| IV. Trial and edit the assessment tool | 11. Analyze feedback from the experiment process. | Tn1. Select the appropriate test sample |
|            |                        | Tn2. Choose the right method of collecting experimental information |
|            |                        | Tn3. Determine the factors affecting the readiness of the subjects participating in the experiment |
|            |                        | Tn4. Eliminate or minimize all factors that influence the readiness of the subjects participating in the experiment |
|            |                        | Pt1. Consider feedback and provide a general comment about the tool |
|            |                        | Pt2. Select some appropriate feedback analysis methods |
|            |                        | Pt3. Separately analyze each type of feedback |
|            |                        | Pt4. Analyze types of feedback with attention to the relationship among them |
|            |                        | Cs1. Consider a number of factors affecting the accuracy of the assessment tool |
|            |                        | Cs2. Determine factors that may affect the accuracy and optimality of the assessment |
|            |                        | Cs3. Modify the assessment tool in order to fix some of those factors |
|            |                        | Cs4. Appropriately edit the assessment tool |

3.4. Pilot test with science pre-service teacher

The pilot aimed to answer the following questions: Is the expression and order of the quality criteria of the model relevant to the reality? Is the model useful for designing competence assessment tool of student? What does the model need to add or adjust?

Sample. The sample consists of 53 third-year pre-service teachers of HNUE, including 23 bilingual pre-service teachers in English. Although there are 11 male pre-service teachers in this sample, this is not a big difference as there are fewer male pre-service teachers in education in Vietnam. The enrollments into HNUE are from the whole country; therefore, there is no regional difference in the sample. These students are trained to teach physics, but they have also been trained in teaching topics that integrate natural subjects, STEM education, and organizing creative experiences.
Tool. To design assessment tasks, we have reviewed 9 studies on standardized teacher assessment tools [32-40]. However, most of these tools are now focused on the contents of the U.S 1990 Standards for Teacher Competence in Educational Assessment of Students. In particular, some tools are questionnaires that require respondents self-assessment; the rest is using objective tests and focuses on knowledge and ethics of assessment so it is not possible to use the tools to collect practical data for the above the CDCAT model.

In the next two weeks, Mr. X wants to assess experimental competence of students of class 10A taught by him, that is the time right after the midterm examination. The class consists of 40 students. The school has only one lab and only 4 sets for each experiment named in the minimum list of schools, the school has 2 cameras and each room is equipped with a projector, speaker, microphone. Design an assessment tool to assess experimental competence of students so that teacher X and a science teacher can organize the assessment activity for two-hour. Describe the process of designing the tool based on the attached questions.

Figure 2: An assessment task to assess CDCAT

Based on the model, we have designed four tools to assess the CDCAT of science pre-service teachers during the learning process of the module "Assessment and Testing in education". Each tool includes assessment tasks and its scoring guide. The descriptions in the scoring guide are specific manifestations of the quality criteria of the model that students may exhibit when performing the assessment tasks. The highest quality criteria of each behavioral indicator are assigned a score of 4, the lower levels are assigned the corresponding scores. Figure 2 is an example of an assessment task using in the learning process of the module. The questions are attached to assessment tasks to evaluate each student's behavior indicator in the CDCAT model.

To ensure the content validity, the assessment tasks and rubric have been reviewed by two assessment experts and two colleagues who were in charge of the same module to consider the suitability of the tasks and rubrics with the model. The assessment experts and subject matter experts have made some small corrections to use the terminology and expressions so that readers understand the author's meaning. Based on the summary of expert opinions, we discussed and revised the tool. Experts have agreed that the tool after editing has ensured the value of the content.

To get the feedback on the model and the assessment tasks developed, we tested on 7 fourth-year science pre-service teachers who have completed the module "Assessment and Evaluation in Education" from the previous semester. To fully test the different levels of competency in the test, the selected students have the test scores of this module ranging from good to moderate. The model was handed over to the students along with the assignments, clarified in the classroom, and the students were encouraged to use them as guidelines to self-assess their work before submission. When pre-service teachers were doing the assessment tasks, we asked them to check if they were unclear or misunderstanding about anything. A few minor verbal corrections were made based on the feedback of these students.

The literature on inter-rater reliability states that for classroom assessments two raters are sufficient to produce acceptable levels of inter-rater agreement. To test the reliability of the tools, in this study two raters independently scored each of the student artifacts. The raters were blind to the other raters and also to the origination of the samples. For items that differ in the two judges' assessments of student achievement, we discuss them with them to edit the scoring. After that, two raters re-evaluate the item on other students’ work. The scoring was edited after discussing the 4 students’ work. The results of two raters mark the work of the remaining three students with this scoring are identical. The sample size in this study is within an acceptable range of 6-12 samples per course and assessment method. Thus, with testing on a small sample, the tool initially guaranteed reliability.
**Data Collection.** The assessment task in figure 2 was assigned to 53 students during the learning process of the module "Assessment and Evaluation in Education". The situation in the assessment task is real and students meet the needs of 10A data collection when needed. The organization steps for these students to do the assessment task are similar to those for 7 students during the testing of the above tool. Students are asked to give feedback on the clarity of the model, the usefulness of the model in guiding students to design the assessment tool. After students submit their works, we give them the scoring guide and asked them to self assess and peer assess. We also graded these works and compared them with students' grading results.

**Data Analysis** and discussion. Comparing the evidence obtained with the model's quality criteria shows that the quality criteria are relevant to the reality and the model has covered the manifold manifestations of this competence. Analyzing the student's performance when experimenting demonstrated the presence of the quality criteria proposed above. For each behavior indicator, when a student met high-quality criteria, he/she also met lower quality criteria. Thus, the order of the quality criteria for each behavioral indicator is also confirmed empirically. The points between us and students are 94% consistency, so we can see that the tool used in this pilot study is reliable. Because of the small sample, scores of students were analyzed using the Guttman analysis [8]. The results of Guttman's analysis have shown that for each behavior indicator, the quality criteria at high levels in the CDCAT model are more difficult than that of lower levels. Thus, the position of the quality criteria of each behavioral indicator has been initially confirmed to be appropriate.

![Figure 3: Equipment for generator, Springs, Sparkling marbles](image)

All 53 students completed the feedback forms. 91% of the students found the model to be easy to understand and use. All the students said they used the model to serve as guidelines before starting their assignment and have a guiding effect for them during the test. Five of the students mentioned the need for discussing the model in the classroom upgraded with anchor papers to help understand the differences in levels better of behavior indicator number 7. Therefore, we chose some student's works in this pilot to make anchor papers. Those works are presented below.

Evidence for St1: Students can only draft an assessment task of conduct experiments and process data in way they are usually assessed, but not in accordance with time conditions and facilities of the situation to be assessed: “conduct experiments and process the data according to the plan presented in the unit: Measurement of friction coefficients in tenth grade textbook”.

Evidence for St2: Students prepared an assessment task of design experimental plan, so students cannot assess pupils at different levels: “1. when observing a buoy bobbing on the surface of the water, Y has an idea of generating electricity from water waves. Using the equipment in Figure 3, design an experimental investigation to help Y to create a simple generator that uses water wave energy”. Students prepare two separate assessment tasks to conduct experiments and process data, so it is suitable for the following situations: "2. Within 10 minutes, please arrange, install and conduct experiments in the unit "Measurement of friction coefficient in tenth-grade textbook; 3. For the data table of the unit "Measure the friction coefficient in tenth-grade textbook" following, please process the data and state the possible causes of the experimental errors and suggest a method to reduce errors."
Evidence for St3: Students developed a number of assessment tasks of experimental investigation that have different difficulty, so it is possible to evaluate students at different levels: “1. Figure 4 is a black spring (in the middle) that has known hardness and other springs. Describe experimental design to determine the hardness of the remaining springs if using only one straight ruler with the smallest division of 1 mm; 2. Give a force meter with a measurement limit of 10N, a monofilament, and a heavyweight of 11-20N. How to determine the approximate weight of the object without damaging the force meter?”. Students have developed assessment tasks of conduct experiments and process data in accordance with the situation and allow pupils to choose their way of expressing: “Within 45 minutes, conduct experiments to measure an indirect quantity of physics. You have the option of a lab kit or an experiment you made yourself in STEM activity”.

Evidence for St4: In addition to the implementation of the st3 level, students also mentioned how to think: “When observing the marbles as shown in Figure 5, I ask myself: how are those marbles made? What factors need to change to make more sparkling marbles? Since then, I have drafted the following assessment task: how do you turn transparent marbles into sparkling marbles if you use a refrigerator, a stove, a water jug, and two pots?”. Students drew rules for themselves in creating ideas for drafted assessment tasks for the next time: “Observing the objects and thinking: how to made the objects, which parameters of the manufacturing need to change to made an object that has a more powerful feature”.

Thus, the model proposed in this study is relevant to reality and has guiding effect for designing assessment tools in science teaching. For some students, it is possible to use students’ work gained in this pilot as anchor papers to help them understand the model before doing another assessment task.

4. Conclusion
One of the difficulties of STEM teaching today is to assess the student's competency. Research results have opened up prospects for future teachers to perform well the design of assessment tools to assess competency in stem teaching.

The content of the model was based on the common and reasonable points of literature discussed earlier and analysis cognitive process, therefore the model will match with teacher education in other countries. The study proposed a model of CDCAT following a multi-step process. Analyzing the literature and the data and double implementing of expert panel have ensured the content validity of the CDCAT model. The internal structure validity of the model is shown by analyzing the results of the survey and the experimental results. Overall, this study suggested that the model is reliable and valid.

The model CDCAT was piloted with a small amount of Vietnamese science pre-service teachers. The research should continue to standardize the model of CDCAT and build the developmental framework by measuring CDCAT on large numbers of pre-service teachers in other countries. The model CDCAT was piloted with small amount of Vietnamese science pre service teachers. The research should continue to standardize the model of CDCAT and build developmental framework by measuring CDCAT on large numbers of pre-service teachers in other country.

5. References
[1] Black P and Wiliam D 1998 Assessment and classroom learning Assessment in Education: Principles, Policy, and Practice 5 p 7–74.
[2] Bol L et al 1998 Influence of experience, grade level, and subject area on teachers’ assessment practices The Journal of Educational Research 91 p323–330.
[3] Stiggins R.J. and Conklin N F 1992 In teachers’ hands: Investigating the practices of classroom assessment Albany: State University of New York Press.
[4] Wiggins G 1989 A true test: Toward more authentic and equitable assessment. Phi Delta Kappan 70 p703–713.
[5] DeLuca C and Klinger D A 2010 Assessment literacy development: Identifying gaps in teacher candidates’ learning Assessment in Education: Principles, Policy & Practice 17 p419–438. doi:10. 1080/0969594X.2010.516643
[6] Volante L and Fazio X 2007 Exploring teacher candidates’ assessment literacy: Implications for teacher education reform and professional development. Canadian Journal of Education 30 p749–770.

[7] Linh Ng and Tra D H 2016 From surveying reality to proposing some solutions to develop competence of designing assessment tool in training and improving improving physics teacher HNUE journal of science 8B p. 213-225

[8] Griffins P 2015 Assessment for teaching Cambridge University Press p. 125–156.

[9] New Zealand Teachers Council 2008 Graduating teacher standards. Retrieved from http://www.teacherscouncil.co.nz/.

[10] Department of Education-Australia 2012 Australian professional standards for teachers. Retrieved from http://www.teacherstandards.aitsl.edu.au/

[11] Department for Education-United Kingdom 2012 Teachers’ standards. Retrieved from http://www.education.gov.uk/schools/teachingandlearning/reviewofstandards/a00205581/teachers-standards1-sep-2012.

[12] Association for Educational Assessment-Europe 2012 European framework of standards for educational assessment 1.0. Rome: Edizioni Nuova Cultura.

[13] American Federation of Teachers, National Council on Measurement in Education, & National Education Association 1990 Standards for teacher competence in educational assessment of students Washington: National Council on Measurement in Education.

[14] America Educational Research Association, American Psychological Association, & National Council on Measurement in Education 1999 Standards for Educational and Psychological Testing. Retrieved from http://www.teststandards.org/.

[15] Assessment Reform Group 2008. Changing assessment practices: Process, principles and standards. Retrieved from http://www.aai.org.uk/content/uploads/2010/06

[16] Black P and Wiliam D 1998 Inside the black box: raising standards through classroom assessment. Granada Learning.

[17] Interstate Teacher Assessment and Support Consortium 2011 InTASC model core teaching standards: a resource for state dialogue. Washington: Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Resources/Publications/InTASC.html.

[18] Joint Committee for Standards on Educational Evaluatio 2015 Classroom assessment standards: practices for PK-12 teachers. Retrieved from http://www.teach.purdue.edu/pcc/JCSE_Assessment_Standards.pdf.

[19] National Board for Professional Teaching Standards 2001 What teachers should know and be able to do. Retrieved from http://www.nbpts.org/what_teachers_should_know.pdf.

[20] National Council for the Accreditation of Teacher Education 2008 Professional standards for the accreditation of teacher preparation institutions. Retrieved from: http://www.ncate.org/public/standards.asp.

[21] Stiggins R 2002. Assessment crisis: The absence of assessment for learning. Phi Delta Kappan 83, no. 10: (pp.758–65).

[22] North Central Regional Educational Laboratory, (n.d.). Indicator: Assessment. Retrieved July 24, 2016, from http://www.ncrel.org/engauge/framework/pro/literacy/prolitin.htm

[23] Stiggins R 1987 Design and development of performance assessment. Educational Measurement: Issues and Practice p33-8

[24] Brualdi A 1998 Implementing performance assessment in the classroom. Practical Assessment, Research & Evaluation 6(2) Retrieved from http://PAREonline.net/getvn.asp?v=6&n=2.

[25] Freeman R and Lewis R 1998 Planning and implementing assessment. London: Kogan Page.

[26] Messick S 1994 The interplay of evidence and consequences in the validation of performance assessments Educational Researcher 23 p13–23.

[27] Moskal B M 2003 Recommendations for developing classroom performance assessments and scoring rubrics Practical Assessment, Research & Evaluation 8(14) Retrieved from http://PAREonline.net/getvn.asp?v=8&n=14

[28] Wiggins G and MC Tighe J 2005 Understanding by design Association for Supervision and Curriculum Development.
[29] Tillema H et al 2011 Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning – A review of research studies Studies in Educational Evaluation 37 p 25–34
[30] Austria Goverment Department of Education, Employment and workplace Relations 2012 TAEASS502B Design and develop assessment tool Retrieved from https://training.gov.au/TrainingComponentFiles/TAE10/TAEASS502B_R1.pdf
[31] Gulikers J et al 2005 The surplus value of an authentic learning environment Computers in Human Behavior 21 p509–521.
[32] Kershaw IV I 1993 Ohio vocational education teachers’ perceived use of student assessment information in educational decision-making. Ohio State University.
[33] Zhang Z and Burry-stock JA 1997 Assessment practices inventory: a multivariate analysis of teachers’ perceived assessment competency. Annual meeting of the American Educational Research Association,Chicago.
[34] Jarr KA 2012 Education practitioners’ interpretation and use of assessment results. University of Iowa. Retrieved from http://ir.uiowa.edu/etd/3317.
[35] Plake B et al 1993. Assessment competencies of teachers: a national survey Educational Measurement: Issues and Practice 12(4) p10–39. doi:10.1111/j.1745-3992.1993.tb00548.x.
[36] Daniel L G and King D A 1998 Knowledge and use of testing and measurement literacy of elementary and secondary teachers. The Journal of Educational Research 91(6) p331–344.
[37] Campbell C et al 2002 Psychometric analysis of an assessment literacy instrument: applicability to pre-service teachers Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus.
[38] Mertler C A 2003 Pre-service versus in-service teachers’ assessment literacy: does classroom experience make a difference? Annual meeting of the Mid-Western Educational Research Association, Columbus.
[39] Christopher DeLuca et al 2016 Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy Educational Assessment 21(4)p248-266, DOI: 10.1080/10627197.2016.1236677
[40] Alonzo D A 2016 Development and Application of a Teacher Assessment for Learning Literacy Tool A thesis in fulfilment of the requirements for the degree of Doctor of Philosophy School of Education Arts and Social Sciences
[41] Biggs J B and Collis K F 1982 Evaluating the quality of learning: The SOLO taxonomy. New York: Academic Press.
[42] Reddy Y M 2011 Design and development of rubrics to improve assessment outcomes Quality Assurance in Education 19 (1) p84-104
[43] Moskal B M and Leydens J A 2000 Scoring rubric development: validity and reliability, Practical Assessment, Research & Evaluation Vol. 7 p71-81,
[44] Sutaphan, S. and Yuenyong, C. (2019). STEM Education Teaching approach: Inquiry from the Context Based. Journal of Physics: Conference Series, 1340 (1), 012003
[45] Chomphuphra P, Chaipidech P, & Yuenyong C 2019 Trends and Research Issues of STEM Education: A Review of Academic Publications from 2007 to 2017. Journal of Physics: Conference Series, 1340 (1), 012069
[46] Duc NM, Linh NQ, &Yuenyong C 2019 Implement of STEM education in Vietnamese high school: unit of acid-base reagent from purple cabbage. Journal of Physics: Conference Series, 1340 (1), 012029