



The Development and Validation of a Mechanical Critical Thinking Scale for High School Students

Kuang-Chao Yu

National Taiwan Normal University, TAIWAN

Kuen-Yi Lin

National Taiwan Normal University, TAIWAN

Shu-Fen Chang

Taipei Municipal Nangang Vocational High School, TAIWAN

Received 30 September 2015 • Revised 3 March 2016 • Accepted 14 April 2016

ABSTRACT

The purpose of this study was to develop a mechanical critical thinking scale for high school students. A stratified random sampling method was used to establish the norms. After pre-tests and item analysis, the scale was determined to have five subtest sections (i.e., recognition of assumptions, induction, deduction, interpretation, and evaluation of arguments) consisting of 25 items. In addition, using 1,954 high school students as participants, this study established a norm and investigated the differences among genders and ages. The results showed that male students had better mechanical critical thinking skills than did female students in terms of induction aptitude, while female students were better in evaluation aptitude. In addition, the mechanical critical thinking performance of 12th-grade students was superior to that of 10th- and 11th-grade students.

Keywords: critical thinking scale, high school students, mechanical, norm

INTRODUCTION

It is important to cultivate high students' interest and skills in engineering in pre-college engineering programs. After investigating the learning process of high school students in engineering design activities, Yu, Lin, and Fan (2013) found that students showed good performances in material application and modeling design, yet their abilities in mechanical designing and functional performance still needed to be improved. However, in order to cultivate students' research and innovation ability in engineering, pre-college engineering programs should focus on not only material application and modeling design but also the improvement of students' abilities in mechanical design and functional performance. For example, traditionally, mechanical designers in engineering always perform mechanical design based on their intuition, creativity, and experience (Tsai, 2001). However, due to the lack of complete work experience and learning opportunity, high school students are

© **Authors.** Terms and conditions of Creative Commons Attribution 4.0 International (CC BY 4.0) apply.

Correspondence: Kuen-Yi Lin, *National Taiwan Normal University, Taiwan.*

✉ linkuenyi@ntnu.edu.tw

State of the literature

- Current research focused on developing general critical thinking scale, but lack of developing domain-specific measurement of critical thinking.
- Current research offers a large number of references about students' performances in general critical thinking skills, but lack of exploring students' performances in domain-specific critical thinking skills.

Contribution of this paper to the literature

- This paper develops a mechanical critical thinking scale for high school students.
- This paper explores high school students' performances in mechanical critical thinking skills.
- This paper explores the differences in performances between students of different genders and ages in the mechanical critical thinking skills.

experiencing difficulties in properly demonstrating their mechanical design abilities in engineering design activities.

Regarding the cultivation of mechanical design ability, Abbas, Choudhary, and Khan (2013) explored the ability requirements of design engineers (including mechanical design engineers) and pointed out that design engineers should possess seven abilities, among which critical thinking is one of the important for cognitive skills. Critical thinking is a complex cognitive process involving the interactions between the learner's knowledge, intention, and skills and the surrounding scenario; therefore, the learner has to establish a series of effective and reasonable judging criteria to clarify and estimate problems for decision making and problem solving (Yeh, Yeh, & Hsieh, 2000). Hence, if high school students have the critical thinking skills for mechanical design, they will be able to properly use their knowledge, intention, and skills to clarify and estimate problems and further to design proper mechanisms to solve problems when involved in engineering design activities.

Existing scales related to critical thinking skills mostly evaluate only general critical thinking skills (Ennis, 1993; Ennis, Millman, & Tomko, 1985; Ennis & Norris, 1989; Facione & Facione, 1992; Watson & Glaser, 1964; Watson & Glaser, 1980). However, Ennis (1989) found that critical thinking has subject specificity (similar to domain specificity); therefore, evaluation scales for critical thinking abilities vary among different fields and require different design considerations to develop more proper educational or evaluation tools (Tiruneh, Cock, Weldelessie, Elen, & Janssen, 2016). A study of Renaud and Murray (2008) using psychology as the learning content showed results consistent with Ennis' opinion. They compared two evaluation tools: domain-specific measure of critical thinking and general critical thinking scale. The results revealed that, using the former tool for evaluation, students in the experimental group would have better performance in critical thinking than those in the control group; using the latter tool, no significant difference was observed between the two groups. For this reason, this study mainly focused on the field of mechanisms and attempted to develop a mechanical critical thinking scale suitable for high school students by integrating scholars' definitions of critical thinking. Furthermore, we established a norm as a reference for

education and evaluation of mechanical design in the future. Specifically, this study aimed to address the following questions: (1) Which content is suitable and should be included in the mechanical critical thinking scale for high school students? (2) Are there any differences in performance between students of different genders in the mechanical critical thinking? (3) Are there any differences in performance between students of different ages in the mechanical critical thinking?

BACKGROUND

Critical thinking skills in Mechanism

Many researchers proposed different views regarding the definition of critical thinking. For example, some considered critical thinking as skills (Halpern, 1997; Paul, 1990), some as processes (Chaffee, 1990; Facione, Sanchez, Facione, & Gainen, 1995), and some as procedures (Bailin, Case, Coombs, & Daniels, 1999; Marzano, Brandt, Hughes, Jones, Presseisen, Rankin, & Suhor, 1988). In summarizing these three notions, critical thinking can be considered a complex cognitive process involving the interactions between the learner's knowledge, intention, and skills and the surrounding scenario. Thus, the learner has to establish a series of effective and reasonable judging criteria to clarify and estimate problems for decision making and problem solving (Yeh, Yeh, & Hsieh, 2000).

A number of researchers developed tools to effectively evaluate learners' critical thinking ability from the aspect of skills. For example, Michelli, Pines, and Oxman-Michelli (1990) and Ennis and Norris (1989) all believed that critical thinking involves many cognitive and meta-cognitive skills, including analysis, interpretation, induction, explanation, evaluation, creative use of information, conclusion drawing, interaction with others, and self-regulation. Nevertheless, the Cornell Critical Thinking Test (Level X) mainly measures four skills: induction, credibility, deduction, and identification of assumptions. The Cornell Critical Thinking Test (Level Z) mainly measures seven skills: induction, credibility, prediction in planning experiments, semantics, deduction, definition, and identification of assumptions (Ennis, Millman, & Tomko, 1985). The Watson-Glaser Critical Thinking Appraisal mainly measures five skills: interpretation, deduction, recognition of assumptions, induction, and evaluation of arguments (Watson & Glaser, 1980).

By summarizing the notions above on the development of critical thinking tests, Yeh, Yeh, and Hsieh (2000) proposed that a critical thinking test should measure the following five skills: (1) Recognition of assumptions: the ability to recognize hidden premises or default stands considered merited in statements; (2) Induction: the ability to induce the most possible results from known information; (3) Deduction: the ability to identify hidden relationships between statements from known statements or premises and to determine from the known premises whether or not this induction is, indeed, the hidden or inevitable result; (4) Interpretation: the ability to find evidence from statements and evaluate the possibility of the induction; (5) Evaluation or arguments: the ability to evaluate the support level of arguments in a question. In addition, Yeh, Yeh, and Hsieh (2000) found that, among the five skills, high

school students had the best performance in recognition of assumption ($M = 3.51$, $SD = 0.98$), followed by deduction ($M = 3.29$, $SD = 1.34$), evaluation of arguments ($M = 2.96$, $SD = 1.18$), induction ($M = 2.87$, $SD = 1.17$), and interpretation ($M = 2.68$, $SD = 1.24$).

Nevertheless, critical thinking should be based on domain-specific knowledge, because one has to apply domain-specific knowledge as well as experience and common sense to perform critical thinking effectively, which will not happen in a “vacuum” condition (Norris, 1985). Ennis (1989) also found that critical thinking had domain specificity. Therefore, critical thinking abilities required for different domains vary, and their evaluation tests vary accordingly, thus requiring different design considerations to develop more ideal educational or evaluation tools. According to the study conducted by Renaud and Murray (2008), domain-specific measurement of critical thinking skills is indeed more helpful when exploring students' critical thinking performances in different domains. For this reason, this study aimed to develop a critical thinking scale in mechanism, in which the five skills summarized by Yeh, Yeh, and Hsieh (2000) were used as major measurement items. This scale can be an effective tool to evaluate mechanical critical thinking skills. In addition to validating the difficulty and discrimination indexes of this tool, a norm was established in this study for future reference.

Gender

Many researchers used gender as a variable when exploring differences in critical thinking. However, the literature review and analysis in this study revealed that most researchers believed that the performances in critical thinking were not significantly different between students of different genders. For example, Yeh, Yeh, and Hsieh (2000) developed a localized critical thinking scale suitable for secondary school and elementary school students and found the critical thinking skills between boys and girls were not significantly different. Terry and Ervin (2012) analyzed college students' performance on the California Critical Thinking Skills Test (CCTST) and stated that gender was not an important factor affecting students' performance on critical thinking. Bensley and Spero (2014) focused on the cultivation of student's critical thinking skills through argument analysis, critical reading, and other methods and found these methods to be helpful in improving student's critical thinking skills and that the difference in performance on the critical reading test (CRT) between students of different genders was not significant. Özyurt (2015) tried to explore computer engineering students' performances in critical thinking disposition, and the results showed that thinking disposition levels of the students did not vary statistically significantly by gender and grade.

However, other researchers pointed out those students of different genders have significant differences in performance on critical thinking. For example, Liu (2010) established a critical thinking scale for high school students including four dimensions: reviewing, reassessment, contemplation, and reflection. Besides, Liu (2010) found out that male students demonstrated significantly stronger analysis, reflection abilities, and overall performance than female students. Lowrie and Diezmann (2011) focused on studying the performances of students of different genders in graphical problem-solving tasks (refers to the problems related

to the tools of Graphical Languages in Mathematics, including number lines, graphs, maps, and diagrams). They found that boys had better performance in graphical tasks than did girls; in particular, when facing graphical tasks of two and three-dimensional representations, boys demonstrated better skills in interpretation of information than did girls. Using a critical thinking scale developed by Halpern (2008), it was found that boys had significantly better performance on the Inference Analysis Test (IAT) than did girls; however, their performances on the Argument Analysis Test (AAT) were not significantly different.

Among these related studies, most of those that revealed an insignificant difference in the performance on critical thinking between students of different genders used general critical thinking tests (Terry & Ervin, 2012; Yeh, Yeh, & Hsieh, 2000). The studies that found a significant difference between students of different genders all used domain-specific critical thinking tests, such as the graphical task used by Lowrie and Diezmann (2011), which is different from traditional critical thinking tests. For this reason, because this study used the mechanical critical thinking scale as an evaluation tool, the performances of different genders tested by this tool were expected to be significantly different.

Age

Very few studies related to critical thinking focused on difference analysis or exploration among different age populations. The literal review and analysis in this study revealed that there are different opinions on whether students of different ages perform differently on critical thinking. For example, the study conducted by Yeh, Yeh, and Hsieh (2000) testing critical thinking skills of secondary and elementary school students found that their critical thinking skills had an uptrend with age; in other words, older students had better performance on critical thinking tests. Although Yeh, Yeh, and Hsieh's study (2000) stated that age was an important influencing factor for critical thinking performance, the study conducted by Terry and Ervin (2012) drew an opposite conclusion according to results of the CCTST carried out in university students. According to Terry and Ervin's study (2012), in university students, older students did not necessarily have better performance on critical thinking.

Because both of the above studies used general critical thinking tests, the type of evaluation tool should not be the primary cause for the difference in opinion. The more probable reason might be the fact that Yeh, Yeh, and Hsieh (2000) mainly studied secondary school students while Terry and Ervin (2012) mainly studied university students. When students grew to university ages, the differences in their performance on critical thinking might become insignificant. On the contrary, before university, age might be a factor involved in the differences in their performance on critical thinking. Because this study used high school students as study participants, the performance on mechanical critical thinking should show statistically significant difference among students of different ages due to difference in maturity.

METHODS

Target Sample

This study was performed in two stages: pre-test and formal test. The purpose of the pre-test was to ensure good reliability and validity of the critical thinking scale developed in this study. Meanwhile, the purpose of the formal test was to establish a norm and perform difference analyses of gender and age. The pre-test was conducted with 137 students, which are selected by purposive sampling, from three high schools in Taipei (including 10th and 11th grades) and one vocational high school (including 11th grade). As for the formal test was conducted with 2,691 students, which are selected by a stratified random sampling method, from eight high schools in the northern region (Keelung, New Taipei, Taipei, and Taoyuan County), four in the central region (Miaoli County, Taichung, Changhua County, and Yunlin County), six in the southern region (Tainan, Kaohsiung, and Pingtung County), and two in the eastern region (Taitung County and Kinmen County). After excluding invalid questionnaires that were incomplete or not returned, 1,954 valid questionnaires (1,131 boys and 823 girls) were recovered, including 766 10th-grade students, 459 11th-grade students, and 729 12th-grade students.

Instrument

The major instrument in this study is the Mechanical Critical Thinking Scale (MCTS) for high school students, which is included a self-developed basic information questionnaire consisting of five items (school name, grade, class, student identification number, and gender) and five mechanical critical thinking skills to be measured (recognition of assumptions, induction, deduction, interpretation, and evaluation of arguments). A higher score indicated a stronger mechanical critical thinking ability. The test time (including instruction time) was 20 minutes. To explain the form of the test, there was a paragraph describing mechanical critical thinking skills and instructing on how to answer the question before each subtest section, followed by an example and introduction.

Taking recognition of assumptions as an example, as shown in Example 1, A is "false," because there is a "pre-existing" idea (i.e., parts is a factor that affects function design) when we state the design conditions. Hence, John must have a "pre-existing" idea in his mind: "The parts I select will affect the function of the doll, and the selection should not take into account cost. B is "true," because functional performance is an important outcome resulting from the doll design, while the combination of parts affects doll function. Therefore, John must have taken "function" as the prior consideration before saying "I have to know the functional conditions for the mechanical doll design before selecting proper parts." C is "false" as well, because parts are functional operation-oriented, and the doll's appearance can be achieved by decoration design and production. Hence, John's words clearly state that the selection of parts does not take appearance as a main consideration.

[Example 1]

John says: I have to know the functional conditions for the mechanical doll design before selecting proper parts.

(A) The conditions are used to limit function, while the main consideration for part selection is cost.

(B) The selection of parts is limited by functional conditions.

(C) The selection of parts is the main factor that affects the doll's appearance.

For the reliability of the MCTS, the split-half method is utilized in analyzing the reliability. The correlation of the sum of item 1 to 3 in five subscales and the sum of item 4 to 5 in five subscales is calculated, and the correlation coefficient is 0.57 ($p < .0001$), indicating moderate levels of correlation. As for the validity of the MCTS, the criterion-related validity is utilized in analyzing the validity. The correlation of students' performance in MCTS and "The Test of Critical-thinking Skills for Primary and Secondary School Students (Yeh, Yeh, & Hsieh, 2000)" is calculated, and the correlation coefficient is 0.13 ($p = .038$), indicating low levels of correlation. Collectively, these psychometric results suggest that the MCTS was of good quality for this low-risk study.

Procedure

Drafting of question items

After the literature review, researchers independently collected the possible items related to senior high school students' mechanical design. From October to December of 2012, two university professors and two senior or vocational high school teachers spent three hours per week together to discuss and draft the test items. Targeting the five mechanical critical thinking skills to be measured (recognition of assumptions, induction, deduction, interpretation, and evaluation of arguments), researchers first determined the form of questions and took into account whether the item topic is close to senior high school students' life experiences for item selection. The selection of item topics was based on current events and topics that the participants were more familiar with. The item statements were drafted corresponding to senior high school students' language and comprehension abilities. Subsequently, three senior or vocational high school teachers were invited to evaluate the suitability of item statements of the questionnaire draft. Twenty senior high school students were asked to provide suggestions for improvement after participating in a test using the draft. Finally, a preliminary scale for pre-test was completed after necessary corrections and modifications.

Pre-tests

The tools used in the pre-test of this study included a self-developed basic information questionnaire consisting of five items (school name, grade, class, student identification

number, and gender) and the preliminary Mechanical Critical Thinking Scale (MCTS) for high school students." In the formal test, the MCTS was used. This scale mainly consisted of five sections (subtests) including recognition of assumptions, induction, deduction, interpretation, and evaluation of arguments. The preliminary version contained five items (questions) in each section, adding up to 25 questions in total. From October 2012 to January 2013, pre-tests were conducted with 137 high school students. During the pre-tests using the preliminary MCTS, the researchers first spent five minutes for instruction and then asked the students to do the questionnaire. The total time was limited to within 20 minutes.

Item analysis and modification based on pre-test results

After recovering the pre-test data, the difficulty and discrimination levels of the scale were immediately analyzed. Based on the analysis results, improper items were deleted or modified or were replaced by more suitable items.

Item analysis after modification

After the pre-test, five questions were revised, the modified scale continued to have 25 questions with five in each section. Each correct answer was given 1 point, and an incorrect answer was given 0 points, giving a full score of 25 points without taking away points for the incorrect answers. To ensure good reliability and validity of the MCTS after modification of certain items, the included items were re-analyzed, and the correlations between subtests were detected to establish a complete version of the MCTS.

Formal tests

After confirmation of reliability and validity of the final MCTS, formal tests in 2,691 senior high school students were conducted between August and October 2013. A total of 1,954 valid questionnaires were successfully recovered.

Norm establishment and difference analyses of gender and age

Based on the recovered questionnaire data, norms for grades 10–12 were established. The effects of gender and age on senior high school students' mechanical critical thinking skills were analyzed to further verify the validity of the MCTS.

Data analysis

In the pre-test stage, item analysis, internal consistency analysis, and Pearson's product-moment correlation analysis were conducted. In the formal test stage, the data analysis methods included descriptive statistics, independent-sample t-test, single-factor analysis of variance (ANOVA), and Scheffe's post-hoc comparison.

Table 1. Item distribution in the revised MCTS

	Recognition of Assumptions	Induction	Deduction	Interpretation	Interpretation	Total
Dimensions	I	II	III	IV	V	5
Items	1–5	6–10	11–15	16–20	21–25	25

RESULTS

Preliminary examination

Item analysis and modification

Item analysis of the preliminary MCTS in this study showed that the discrimination indexes were between 0.14 and 0.57 with an average of 0.39 and the difficulty indexes were between 0.34 and 0.88 with an average of 0.67. Ebel believed that an item should be removed if its discrimination index (D) is below 0.19 and is acceptable but needs to be modified if D is between 0.20 and 0.29, and the optimal difficulty index (P) is between 0.40 and 0.80 (Guo, 1996). According to Ebel's standards, modifications or deletions were performed mainly based on the discrimination index and difficulty index derived from item analysis of the pre-test scale. Modifications were performed for items 1 (P = 0.86, D = 0.22), 4 (P = 0.88), and 5 (P = 0.84, D = 0.24) in the subtest section of "recognition of assumptions"; items 12 (P = 0.84) and 15 (D = 0.27) in the subtest section of "deduction"; and items 16 (P = 0.84) and 20 (P = 0.83) in "interpretation." Items 23 (D = 0.16) and 24 (D = 0.14) in the subtest of "evaluations of arguments" were deleted and replaced with new items designed after discussion with specialists. In summary, seven items (1, 4, 5, 12, 15, 16, and 20) were deleted, and two items (23 and 24) were deleted and replaced.

Item analysis of the modified scale

The revised (after modification and deletion) MCTS in this study included 25 items, the distribution of which is shown in **Table 1**. The scoring method after revision was the same as in the pre-test, and the maximum score was still 25.

The discrimination indexes of the revised MCTS in this study were between 0.17 and 0.68 with an average of 0.45, and the difficulty indexes were between 0.37 and 0.90 with an average of 0.61 (details shown in **Table 2**). Therefore, the scale developed in this study consisted of the items with moderate difficulty and satisfactory discrimination.

Table 2. Discrimination and difficulty indexes of the revised MCTS

MCTS	Items	Difficulty Index	Correct Ratio of High-score Group	Correct Ratio of Low-score Group	Discrimination Index
Recognition of Assumptions	1	0.90	0.97	0.76	0.21
	2	0.57	0.75	0.37	0.38
	3	0.71	0.92	0.51	0.41
	4	0.83	0.96	0.58	0.38
	5	0.47	0.59	0.37	0.22
Induction	6	0.54	0.82	0.32	0.50
	7	0.60	0.86	0.40	0.46
	8	0.44	0.67	0.31	0.36
	9	0.49	0.71	0.26	0.45
	10	0.64	0.87	0.36	0.51
Deduction	11	0.63	0.83	0.39	0.44
	12	0.72	0.92	0.42	0.50
	13	0.71	0.97	0.33	0.64
	14	0.65	0.94	0.26	0.68
	15	0.37	0.55	0.27	0.28
Interpretation	16	0.81	0.99	0.45	0.54
	17	0.53	0.82	0.23	0.59
	18	0.40	0.50	0.33	0.17
	19	0.37	0.51	0.29	0.22
	20	0.74	0.95	0.41	0.54
Evaluation of Arguments	21	0.60	0.86	0.33	0.53
	22	0.71	0.93	0.41	0.52
	23	0.62	0.88	0.32	0.56
	24	0.49	0.71	0.32	0.39
	25	0.69	0.92	0.36	0.56
Average		0.61	0.82	0.37	0.45

Correlation analysis

Correlation analyses were performed to understand the correlations between the overall performance and the score in each subtest using the MCTS for high school students. As shown in **Table 3**, all subtest scores were significantly correlated with the total score with the correlation coefficients between 0.59 and 0.73 ($p < .0001$), indicating moderate levels of correlation (Chou, 2009). In addition, the correlations between the subtest scores were also significant, with the correlation coefficients being between 0.24 and 0.39 ($p < .0001$), indicating low to moderate levels of correlation (Chou, 2009).

Table 3. Correlations between the total score and each subtest score when using the revised MCTS

MCTS	Recognition of Assumptions	Induction	Deduction	Interpretation	Interpretation	Total
Recognition of Assumptions	-	0.24*	0.30*	0.27*	0.26*	0.59*
Induction		-	0.35*	0.30*	0.29*	0.66*
Deduction			-	0.39*	0.37*	0.73*
Interpretation				-	0.36*	0.68*
Evaluation of Arguments					-	0.70*
Total						-

*p < .001

Table 4. The means and SD of senior high school students' scores when using the MCTS

	Male (n = 455)		Female (n = 311)		Total (N = 766)	
	M	SD	M	SD	M	SD
Grade 10						
Recognition of Assumptions	3.45	1.14	3.27	1.10	3.38	1.12
Induction	2.70	1.33	2.47	1.22	2.61	1.29
Deduction	2.97	1.33	3.02	1.21	2.99	1.28
Interpretation	2.77	1.24	2.72	1.25	2.75	1.24
Evaluation of Arguments	2.78	1.41	3.07	1.28	2.89	1.36
Total	14.67	4.34	14.55	3.97	14.62	4.19
Grade 11	Male (n = 297)		Female (n = 162)		Total (N = 459)	
Recognition of Assumptions	3.25	1.18	3.09	1.28	3.19	1.22
Induction	2.49	1.28	2.35	1.30	2.44	1.29
Deduction	2.69	1.37	2.51	1.33	2.63	1.36
Interpretation	2.54	1.30	2.51	1.28	2.53	1.29
Evaluation of Arguments	2.62	1.35	2.71	1.40	2.65	1.37
Total	13.59	4.26	13.17	4.45	13.44	4.33
Grade 12	Male (n = 379)		Female (n = 350)		Total (N = 729)	
Recognition of Assumptions	3.74	0.97	3.75	0.98	3.74	0.97
Induction	3.17	1.24	2.84	1.24	3.01	1.25
Deduction	3.42	1.23	3.52	1.21	3.47	1.22
Interpretation	3.22	1.13	3.04	1.00	3.14	1.07
Evaluation of Arguments	3.67	1.32	3.63	1.27	3.65	1.30
Total	17.22	3.86	16.78	3.36	17.01	3.63
Grade 10-12	Male (n = 1,131)		Female (n = 823)		Total (N = 1,954)	
Recognition of Assumptions	3.50	1.11	3.44	1.12	3.47	1.12
Induction	2.80	1.32	2.60	1.26	2.72	1.30
Deduction	3.05	1.34	3.13	1.29	3.08	1.32
Interpretation	2.86	1.25	2.82	1.17	2.84	1.22
Evaluation of Arguments	3.03	1.44	3.24	1.35	3.12	1.40
Total	15.24	4.41	15.23	4.08	15.24	4.27

Table 5. t-tests of high school students' scores in the tests using the proposed MCTS

MCTS	Male (N = 1,131)		Female (N = 823)		t-value	p-value
	M	SD	M	SD		
Recognition of Assumptions	3.50	1.11	3.44	1.12	1.13	0.26
Induction	2.80	1.32	2.60	1.26	3.44*	0.001
Deduction	3.05	1.34	3.13	1.29	-1.47	0.14
Interpretation	2.86	1.25	2.82	1.17	0.83	0.41
Evaluation of Arguments	3.03	1.44	3.24	1.35	-3.17*	0.002
Total	15.24	4.41	15.23	4.08	0.09	0.93

* $p < .05$

Norms

This study enrolled 1,954 high school students in the tests, including 766 10th-grade students, 459 11th-grade students, and 729 12th-grade students, to establish norms using the proposed MCTS. The means and standard deviations (SD) of scores of female and male students from each grade are shown in **Table 4**. Results showed that, when using the proposed scale, the 10th graders had an average score of 14.62 ($SD = 4.19$), the 11th graders had 13.44 ($SD = 4.33$), and the 12th graders had 17.01 ($SD = 3.63$) with an overall average score of 15.24 ($SD = 4.27$).

The analysis of differences of gender and age in MCTS

As shown in **Table 5**, the male students' mechanical thinking skills in "induction" were superior to those of female students, showing a statistically significant difference ($t(1952) = 3.44, p = .001$), whereas the female students showed stronger mechanical thinking skills in "evaluation of arguments" than did male students with a statistically significant difference ($t(1832.43) = 3.17, p = .002$).

As shown in **Table 6**, senior high school students of different ages were scored significantly differently in subtests of "recognition of assumptions" ($F(2,1951) = 40.22, p < .001$), "induction" ($F(2,1951) = 32.86, p < .001$), "deduction" ($F(2,1951) = 64.92, p < .001$), "interpretation" ($F(2,1951) = 39.91, p < .001$), "evaluation of arguments" ($F(2,1951) = 95.97, p < .001$), and the overall performance of mechanical critical thinking skills ($F(2,1951) = 125.26, p < .001$). In all six dimensions, 12th-grade students showed superior performances to 10th- and 11th-grade students with statistically significant differences. Other than deduction ability, 10th-grade students showed superior performance to 11th-grade students in all other dimensions (recognition of assumptions, induction, interpretation, evaluation of arguments, and the overall mechanical critical thinking skills) with statistically significant differences.

Table 6. Single-factor ANOVA of test scores using the proposed MCTS at different ages

MCTS	Grade						F-value	Scheffe
	10th-grade (n = 766)		11th-grade (n = 459)		12th-grade (n = 729)			
	M	SD	M	SD	M	SD		
Recognition of Assumptions	3.38	1.12	3.19	1.22	3.74	0.97	40.22*	12th > 10th 12th > 11th 10th > 11th
Induction	2.61	1.29	2.44	1.29	3.01	1.25	32.86*	12th > 10th 12th > 11th 10th > 11th
Deduction	2.99	1.28	2.63	1.36	3.47	1.22	64.92*	12th > 10th 12th > 11th
Interpretation	2.75	1.24	2.53	1.29	3.14	1.07	39.91*	12th > 10th 12th > 11th 10th > 11th
Evaluation of Arguments	2.89	1.36	2.65	1.37	3.65	1.30	95.97*	12th > 10th 12th > 11th 10th > 11th
Total	14.62	4.19	13.44	4.33	17.01	3.63	125.26*	12th > 10th 12th > 11th 10th > 11th

* $p < .001$

DISCUSSION

The main purpose of this study was to develop a scale of mechanical critical thinking skills and to use this tool to establish norms of the mechanical critical thinking skills of high school students in Taiwan. In addition, this study explored the differences in mechanical critical thinking skills between high school students of different genders and ages. It can be seen from the aforementioned analysis results that students had the best performance in recognition of assumptions followed by evaluation of arguments, deduction, interpretation, and induction. This result is slightly different from that of Yeh, Yeh, and Hsieh (2000), who found that secondary school students had the best performance in recognition of assumptions, followed by deduction, evaluation of arguments, induction, and interpretation. When using the mechanical critical thinking and general critical thinking scales, senior high school students showed slightly different performances in "evaluation of arguments and deduction" and "interpretation and induction." This conclusion is consistent with most researchers' opinions (Ennis, 1989; Norris, 1985; Renaud & Murray, 2008), i.e., critical thinking is domain specific. Therefore, the mechanical critical thinking scale developed in this study should help future researchers evaluate high school students' mechanical critical thinking skills and cultivate student's mechanical critical thinking skills more effectively.

Furthermore, from the perspective of the gender variable, our analysis revealed that students of different genders performed significantly differently in mechanical critical

thinking tests, which indicated not only those male students performed superiorly at times but also that female students performed superiorly at times. Regarding induction, we found that boys had superior performance, echoing Halpern's (2008) results showing that boys had better performance in "being able to induce the most possible results from given information." Regarding evaluation of arguments, this study discovered that girls performed superiorly. To our knowledge, the majority of existing studies found that boys had advantages in critical thinking ability, yet no relevant studies pointed out girls' advantages in critical thinking tests. Hence, this study's finding that female students were superior to male students in "performance of being able to evaluate the support level of arguments in a question" is worth exploring further to reveal the underlying reason.

Moreover, from the perspective of the age variable, this study discovered that, in the six dimensions of recognition of assumptions, induction, deduction, interpretation, evaluation of arguments, and the overall mechanical critical thinking ability, 18-year-old students showed superior performances to 16- and 17-year-old students with statistically significant differences. This finding is consistent with the study conducted by Yeh, Yeh, and Hsieh (2000) that older students have stronger critical thinking skills. However, this study also found that, other than deduction ability, 16-year-old students showed superior performance to 17-year-old students in all other dimensions (recognition of assumptions, induction, interpretation, evaluation of arguments, and the overall mechanical critical thinking skills) with statistically significant differences. This contradicts the study of Yeh, Yeh, and Hsieh (2000). Because regional difference between urban and rural areas was already taken into account during sampling in this study, it cannot be used to explain the discrepancy between the two studies. Therefore, the actual reason needs to be explored and clarified in the future.

CONCLUSION AND IMPLICATIONS

In this study, the MCTS was developed through rigorous procedures. This scale has an average discrimination index of 0.45 and an average difficulty index of 0.61, suggesting a moderate difficulty and good discrimination of the items included. Considering that critical thinking has domain specificity (Ennis, 1989; Norris, 1985; Renaud & Murray, 2008), the proposed scale is recommended to researchers working on cultivating high school students' mechanical critical thinking skills. Using this scale, researchers can investigate students' critical thinking performance more effectively and avoid the problem of being unable to properly evaluate the educational achievements due to the use of general critical thinking tests (Renaud & Murray, 2008).

Regarding the gender variable, a major conclusion of this study is that male students have better critical thinking ability than female students do in induction, whereas female students have better critical thinking ability than male students do in evaluation of arguments. Therefore, boys and girls each have their own advantages in mechanical critical thinking skills. In the future, if one wants to improve senior high school students' performance in mechanical design and function based on Yu, Lin, and Fan's (2013) study, the gender difference should be

considered to guide students' learning process to cultivate students' advanced mechanical innovation ability.

Regarding the age variable, a major conclusion of this study is that the mechanical critical thinking ability of 12th-grade students is stronger than that of 10th- and 11th-grade students. However, regardless of the superior performance on mechanical critical thinking, the performances of 12th-grade students in induction and interpretation were all below their overall average level. This finding suggests that future mechanical critical thinking education for senior high school students should emphasize training in induction and interpretation abilities so that the students' mechanical critical thinking skills can be more comprehensive.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of Science and Technology of the Republic of China under Contract numbers NSC101-2511-S003-039 and MOST 104-2628-S-003-001. The findings and recommendations contained in this article are those of the authors and do not necessarily reflect those of the Ministry of Science and Technology. We are extremely grateful to the reviewers for their helpful comments, and the teachers and students who participated in this study.

REFERENCES

- Abbas, A., Choudhary, M. A., & Khan, N. (2013). An assessment of competency requirements of design engineers in Pakistan and development of a framework for improvement. *Technology Management in the IT-Driven Services (PICMET), 2013 Proceedings of PICMET'13*, 1723-1737. (2013, October 26). Retrieved from <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6641698&isnumber=6641565>
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Common misconceptions of critical thinking. *Journal of Curriculum Studies*, 31(3), 269-283.
- Bensley, D. A., & Spero, R. A. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity*, 12, 55-68.
- Chaffee, J. (1990). *Thinking critically* (3rd ed.). Boston: Houghton Mifflin.
- Chou, H. C. (2009). *Quantitative Research and Statistical Analysis – Examples of the Chinese Version of SPSS Data Analysis*. Taipei: Wunan.
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4-10.
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32(3), 179-186.
- Ennis, R. H., Millman, J., & Tomko, T. N. (1985). *Cornell critical thinking tests, level X & level Z-manual*. CA: Midwest Publications.
- Ennis, R. H., & Norris, S. P. (1989). Evaluation critical thinking. *Educational Leadership*, 42(8), 40-45.
- Facione, P. A. & Facione N. C. (1992). *The California critical thinking dispositions inventory*. Millbrae, CA: California Academic Press.
- Facione, P. A., Sanchez, C. A., Facione, N. C., & Gainen, J. (1995). The dispositions toward critical thinking. *The Journal of General Education*, 44(1), 1-25.

- Halpern, D. F. (1997). *Critical thinking across the curriculum: A brief edition of thought and knowledge*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Halpern, D. F. (2008). Is intelligence critical thinking? Why we need a new definition of intelligence. In P. C. Kyllonen, R. D. Roberts, L. Stankov (Eds.), *Extending intelligence. Enhancement and new constructs* (pp. 293–310). New York: Lawrence Erlbaum Associates.
- Kuo, S. Y. (1996). *Psychological and Educational Testing*. Taipei: Jing-Hua Bookstore.
- Liu, H. Y. (2010). The development of the senior high school students' attitude scale on critical thinking. *Journal of Education & Psychology*, 33(3), 139–174.
- Lowrie, T., & Diezmann, C. M. (2011). Solving graphics tasks: Gender differences in middle-school students. *Learning and Instruction*, 21, 109–125.
- Marzano, R. J., Brandt, R. S., Hughes, C. S., Jones, B. F., Presseisen, C. S., Rankin, S. C., & Suhor, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Michelli, N. M., Pines, R., & Oxman-Michelli, W. (1990). *Collaboration for critical thinking in teacher education: The Montclair state college model*. Upper Montclair, NJ: Institute for Critical Thinking.
- Norris, S. P. (1985). Synthesis of research on critical thinking. *Educational Leadership*, 42(8), 40–45.
- Özyurt, Ö. (2015). Examining the critical thinking dispositions and the problem solving skills of computer engineering students. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(2), 353–361.
- Paul, R. W. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world*. Rohnert Park, CA: The Center for Critical Thinking & Moral Critique, Sonoma State University.
- Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, 3, 85–93.
- Terry, N., & Ervin, B. (2012). Student performance on the California critical thinking skills tests. *Academy of Educational Leadership Journal*, 16, 25–34.
- Tiruneh, D. T., Cock, M. D., Weldeslassie, A. G., Elen, J., & Janssen, R. (2016). Measuring critical thinking in physics: Development and validation of a critical thinking test in electricity and magnetism. *International Journal of Science and Mathematics Education*. Advance online publication. doi:10.1007/s10763-016-9723-0
- Tsai, L. W. (2001). *Mechanism design: enumeration of kinematic structures according to function*. Boca Raton, FL: CRC Press.
- Watson, G., & Glaser, E. (1964). *Critical thinking appraisal manual*. NY: Harcourt, Brace, and World.
- Watson, G., & Glaser, E. (1980). *Watson-Glaser critical thinking appraisal manual (Forms A & B)*. Cleveland: Harcourt, Brace, Jovanovich (Psychological Corporation).
- Yeh, Y. C., Yeh, P. L., & Hsieh, C. C. (2000). The development of “the test of critical-thinking skills for primary and secondary school students.” *Psychological Testing*, 47(1), 47–55.
- Yu, K. C., Lin, K. Y., & Fan, S. C. (2013). How high school students apply their knowledge in engineering design projects. *International Journal of Engineering Education*, 29(6), 1–11.

<http://iserjournals.com/journals/eurasia>