Development of an Assessment Toolkit to Measure Higher Order Thinking Skills among Secondary School Learners

Loh Sau Cheong, PhD[1], Loo Chu Kiong, PhD[2], Maria Felicitas M. Mamauag, PhD[3], & Sabrena G. Arosh[4]

[1 4] *Faculty of Education, University of Malaya*

[2] *Faculty of Computer Science and Information Technology, University of Malaya*

[3] *CAREERsense, HELP University*

Abstract: **Higher order thinking skills (HOTS) emerged as the focus of student development across the primary to post-secondary levels since the publication of the Malaysian Education Blueprint (2013-2025). This blueprint outlines the need for thinking skills encompassing creativity, innovation, critical thinking, and logical reasoning for students to adapt within the ever-changing technological landscape of the new millennia (Ministry of Education, 2013). Despite MOE's effort to implement PT3 (Form 3 Assessment) and HOTS questions in national school-leaving examinations, it apparently failed to accurately measure the levels of HOTS independent of content knowledge among Malaysian students. This study, therefore, was conceived to address the need for a more valid and reliable measurement of HOTS contextualized in a Malaysian setting. This theoretical paper is an attempt to set a framework by which HOTS' underlying measurement is anchored on. Based on a review of literature, operational definitions of the constructs on critical thinking, problem-solving, and creativity were derived. Three types of measures were developed, namely: (a) syllogisms, (b) situational judgment test (SJT), and (c) creativity. Thirty (30) syllogisms items to measure critical thinking, while 50 items of situational judgment tests to measure problem-solving were generated. The creativity test is adapted from an existing measure. Translated into Bahasa Malaysia, the first two tests are currently undergoing expert review prior to pilot-testing. These tests will be standardized using modern Item Response Theory (IRT) techniques, like Rasch modeling. The essential steps moving towards the full development of the HOTS assessment toolkit for the millennials in Malaysia are further discussed.**

INTRODUCTION

In preparing students for life in the 21st Century workplace, many nations around the world have taken to examining the worth of higher order thinking skills (HOTS) in education. Recent years have been witness to drastic transformations in the local education system where the Malaysian government took steps to implement the integration of HOTS within the school syllabi and introduced an overhaul of national assessments with an emphasis on these skills to pre-empt the need for these capabilities in school graduates of tomorrow. Changes such as these were in line with the publication of the National Education Blueprint in 2013 which outlined the need for focused assessment of HOTS within the schooling system (Ministry of Education, 2013). However, it could be seen that the immediate integration of HOTS assessment within the national examinations caused much uproar when only 4 896 *Ujian Pencapaian Sekolah Rendah* (Primary School Achievement Examination, UPSR) students managed to attain all As compared to more than 38 000 the previous year (Azura, 2016). A less severe drop was noted among *Sijil Pelajaran Malaysia* (Malaysian Certificate of Education, SPM) candidates which prompted a statement from the Minister of Education, Datuk Seri Mahdzir Khalid recommending the need to analyse student weaknesses in answering HOTS questions (Naj, 2017).

Thus, the need for an assessment tool of HOTS was identified, particularly for secondary school learners who were closer to the completion of the national syllabus and poised to enter the world of work. It was important for the tool to be able to accurately assess the level of HOTS in students outside of the syllabus content as HOTS were expected to be a transferable skill that applies across various contexts. Assessment of HOTS beyond the constraints of curricula would allow for more accurate measurement as it would be uninfluenced by students' knowledge of subject content. This paper will discuss the conceptualization of a toolkit for assessing HOTS among secondary school students. As the project is still in its initial stages, the current discussion will focus on the definition of constructs and test development framework, before providing information about future directions.

LITERATURE REVIEW

Higher order thinking skills (HOTS), although frequently mentioned in literature, are rarely well-defined as individual constructs. Most terminology shows gross overlap among constructs and studies with clear delineations are rare. Budsankom, Sawangboon, Damrongpanit, and Chuensirimongkol (2015) provide an overall definition of HOTS as the "ability or expertise to find answers or achieve target goals through various forms of thinking processes" (p. 2640). However, the term HOTS is often used interchangeably with critical thinking and definitions of critical thinking often incorporate aspects of other HOTS such as evaluation, analysis, and synthesis. Bloom's taxonomy of thinking skills served as a reference point around which the following review of literature was composed, in addition to contemporary constructs considered integral to HOTS.

Critical Thinking

Most studies of critical thinking go hand in hand with some form of classroom intervention aimed to increase the levels of critical thinking in students. For example, Çavdar and Doe (2012) studied the use of guided writing assignments in increasing critical thinking of a group of political science students at Colorado State University. Critical thinking in their study was defined as "a set of strategies to help students develop reflective analysis and evaluation of interpretations or explanations, including one's own, to decide what to believe or what to do" (p. 298, Çavdar & Doe, 2012). The Watson-Glaser Critical Thinking Appraisal (1952) was used as a framework for the writing assignment to build on the skills defined within in.

The Watson-Glaser Critical Thinking Appraisal proposes five skills; inference which refers to "the ability to derive logical conclusions from the premises of varied approaches", recognition of assumptions which is "the ability to recognize assumptions and presuppositions implicit in the approaches", deductions which is "the ability to judge whether propositions made by the approaches can be logically drawn from the evidence", interpretation which is "the ability to judge whether the conclusions and arguments made by the approaches can be logically drawn", and evaluation of arguments which is "the ability to distinguish relevant, strong, and weak arguments" (p. 299, Çavdar & Doe, 2012).

These skills were embedded into a two-part writing assignment; a draft which incorporated the skills of inference, recognition of assumptions, interpretation, and evaluation of arguments, and a final paper that relied heavily on the skill of deductions and provided an opportunity for reflection of learning (Çavdar & Doe, 2012). In the draft, students described two schools of thought, applied it to a real-life scenario or situation, and conducted a comparison of the merits and weakness of both schools of thought. In the final paper, students revised their arguments in light of new information that they would uncover, and a postscript was requested for students to reflect upon their learning through the process of the assignment (Çavdar & Doe, 2012).

Student writing showed a clear improvement across the two papers, with their points becoming more accurate and justified in the final paper which was taken as an indicator that critical thinking was improved by the exercise (Çavdar & Doe, 2012). Students themselves identified three benefits of the assignment in the postscript; the final revision helped them understand the concepts better, they felt their ideas were challenged with the additional information and feedback from instructors, and they acquired greater learning about the implications of the theories used (Çavdar & Doe, 2012).

Another study also based on the Watson-Glaser Critical Thinking Appraisal used the instrument developed to provide quantitative data on students' critical thinking. The study by Barnett and Francis (2012) defined critical thinking as "the ability to assess and apply evidence in order to support or evaluate an argument" (p. 204). A quasi-experimental, pre-test post-test study was designed to test the effects of using quizzes with questions targeted at HOTS where students completed the Watson-Glaser Critical Thinking Appraisal before and after the intervention to get a clearer idea of the development of critical thinking in the five skills outlined above (Barnett & Francis, 2012). 147 Educational Psychology students from Northwest Missouri State

University completed the Watson-Glaser Critical Thinking Appraisal (Short Form) which consisted of 40 multiple choice items across five subscales before being assigned to one of three conditions; quizzes with factual multiple choice questions, quizzes with essay questions which required critical thinking of reading material, or quizzes with factual essays (Barnett & Francis, 2012). Although students who completed the HOTS-based quiz showed no significant difference in critical thinking as measured in Watson-Glaser, they performed significantly better in classroom assessments than the other two groups, and all three groups showed an increase in critical thinking skills across the semester (Barnett & Francis, 2012). Though the Watson-Glaser framework seems to be an appropriate theoretical framework of critical thinking, the instrument itself showed less than satisfactory alpha reliabilities (between .23 to .78 across five subscales) which lead to the researchers of this study only using total scores ($\alpha = .71$). In this manner, other assessments could be better representations of critical thinking.

Other measures of critical thinking present subject specific assessments. For example, a study by Ramos, Dolipas, and Villamor (2013) examined the relationship between HOTS and academic performance in university Physics majors. 393 students from a General Physics class in first year at Benguet State University were administered a test consisting of 60 multiple-choice questions covering topics from the class (Ramos et al., 2013). The test was formulated by the teacher with HOTS distribution done in accordance with the Quellmalz taxonomy, and a five-point Likert scale was applied to the final scores to determine their overall level of HOTS (Ramos et al., 2013). Academic performance was measured in the students' final grades for the class (Ramos et al., 2013). In this study, the term HOTS was assumed to be interchangeable with critical thinking which was defined in terms of four cognitive processes; inference, analysis, evaluation, and comparison respectively (Ramos et al., 2013).

The study concluded that students in general demonstrated below average performance in levels of HOTS, and there were no significant differences between males in females (Ramos et al., 2013). Between genders, proficiency in different areas lead to better academic performance; males needed to be better at analysis, comparison and evaluation for better academic performance, whereas females needed to be better at analysis, inference, and evaluation (Ramos et al., 2013). However, the authors failed to elaborate how the items were related to each level of the Quellmalz taxonomy, and construct validity was not clearly explained. In this manner, the construction of the assessment could not be clearly understood.

Another subject-specific study was conducted by Pardamean (2012) at the University of Southern California with a cross-sectional sample of 98 dental students from first- through third-year of study. Pardamean (2012) sought to examine the degree of change in critical thinking skills as the students progressed through the problem-based learning classes. Critical thinking in this study was defined as "the process of determining the authenticity, accuracy, and worth of information or knowledge claims" (p.443, Pardamean, 2012) and students were administered the Health Sciences Reasoning Test (HSRT) to measure the levels of critical thinking. The HSRT is a standardised multiple-choice test with 33 items across five subscales; analysis, inference, evaluation, deductive, and inductive reasoning which are related to three core aspects of critical thinking at a college level, namely analysing problems, drawing inference and evaluation of those inferences (Pardamean, 2012). Overall, no significant continuous increase was found in the students' overall critical thinking score, and this finding applied to all four subscales with the exception of inductive reasoning (Pardamean, 2012). There were no significant differences to be found across the three student levels, but some differences existed according to gender, race, education level, and use of English as a first language (Pardamean, 2012).

Where most of the previous studies looked to examine critical thinking within the classroom environment, Butler (2012) sought to study critical thinking and its relation to real world outcomes. Beyond subject-specific measures of critical thinking, the Halpern Critical Thinking Assessment (HCTA) looks at critical thinking in everyday life. The developer of the instrument defines critical thinking as the use of cognitive strategies in order to increase the possibility of attaining the desired outcome, and is related to thinking in a purposeful, logical, and goal-directed manner (Halpern, 2010, as cited in Butler, 2012). In this manner, the HCTA presents 25 everyday scenarios that elicit open-ended and forced-choice responses which are assessed according to five categories of thinking skills; verbal reasoning, argument analysis, using thought to test hypotheses, reasoning based on likelihood or uncertainty, and general decision-making or problem-solving (Butler, 2012). The assessment is entirely computerized from administration to scoring, where the forced-choice responses are immediately scored but the open-ended responses are scored by trained graders with a maximum score on the whole assessment up to 194 (Butler, 2012). Butler (2012) conducted a correlational study with the Real World Outcomes scale, and found that a moderate relationship existed between the measure of critical thinking and the inventory of real world outcomes.

Aside from the outcomes of critical thinking, a meta-analysis done by Budsankom et al. (2015) looked at various factors that affected student HOTS. In their study, critical thinking was defined as "the ability to evaluate and consider things by searching for reliable and sufficient information before making decisions,

solving problems, evaluating situations and taking action on any tasks with the most appropriate and accurate ways" (p. 2641, Budsankom et al., 2015) which can be seen to incorporate elements of analysis, evaluation, and problem-solving. Through the use of structural equation modelling, the researchers found that the main factor that affected HOTS was student psychological characteristics which referred to a combination of personality traits and thinking processes tied to learning strategies (Budsankom et al., 2015). Classroom environment and intellectual characteristics were seen to have effects as well but the main contribution was attributed to psychological characteristics (Budsankom et al., 2015). It was unclear from the meta-analysis how critical thinking was measured but studies that were included in Budsankom et al.'s work fulfilled the criteria that at least two of the constructs from their proposed model were studied, bivariate correlation values were reported, students were from government schools, and sufficient information was available regarding effect sizes of each study leading to a total of 166 studies being included in the meta-analysis.

Several theoretical frameworks relating to critical thinking have also been developed by researchers in reaching a unified definition of the construct. One of the most notable is the KSAVE framework which has been used in curriculum development for 21st Century Skills (Binkley et al., 2012). Within the framework, constructs are broken down according to the requisite knowledge, skills, and attitudes or values. In defining critical thinking, the relevant knowledge involves knowing how to think systematically and evaluate given evidence, and solve problems with clear articulation of thought process and ideas (Binkley et al., 2012). The two main skills identified related to systems thinking and effective reasoning, while the associated values, attitudes, and ethics pertained to the ability to make reasoned judgments, having an attitudinal disposition that leaned toward open-minded and inquisitiveness, and keen problem solving (Binkley et al., 2012). The use of systems or systematic thinking was tied to other skills of analysis, evaluation, interpretation, and synthesis of information (Binkley et al., 2012).

Though no studies examined the framework exclusively, the authors cited several interventions that were currently in place to boost critical thinking in various populations; Primum is an assessment for medical practitioners to practise diagnosis within the boundaries of a fictitious case study, World Class Tests which are computerised tests of mathematics, science, design and technology, and a Virtual Performance Assessment which allows for testing of students' inquiry skills within the context of the field of Life Sciences (Binkley et al., 2012).

A more recent framework was presented by Dwyer, Hogan, and Stewart (2014) that looked to present a holistic, integrated model of critical thinking. In this model, critical thinking was defined as "a metacognitive process, consisting of a number of sub-skills (e.g. analysis, evaluation and inference) that, when used appropriately, increases the chances of producing a logical conclusion to an argument or solution to a problem" (p.43, Dwyer et al., 2014) where it is similar to the definition of Halpern (2010, as cited by Butler, 2012) above in the sense that both definitions see critical thinking as a means of increasing the probability of forming a reasonable, logical solution.

The Dwyer et al. (2014) framework presents critical thinking as being the result of six components; memory, comprehension, analysis, evaluation, inference, and reflective judgment. Memory basically refers to knowledge or information that the other five skills are built on; without the information, the remaining processes cannot function (Dwyer et al., 2014). Comprehension is seen as a level above memory, in which memory consists of schemas of known information stored within long-term memory, and comprehension allows for the modification of these schemas in light of new information (Dwyer et al., 2014). Analysis is classified as a skill which comes into use when constructing arguments; with analysis, the individual is able to identify components of an argument and its respective role within context of the argument including audio and visual cues from the proposer of the argument (Dwyer et al., 2014).

Evaluation is related to the skill of weighing the strengths and weaknesses of an argument, based not only on the strength of the underlying logic, but also the relevance of the claims, and the credibility of the claimant (Dwyer et al., 2014). Inference refers to the process of sourcing further credible, logical, and valid information in order to reach a conclusion; it is said to be similar to synthesis but distinct in that beyond merely gathering information from multiple credible sources, alternative information is created based on the gathered information and all information gathered or created is judged according to its credibility and usefulness in a way that also overlaps with evaluation (Dwyer et al., 2014). Finally, reflective judgment refers to an awareness of the limitations of one's knowledge and its subsequent effects on the individual's reasoning process, which is related to acknowledging that current boundaries of knowledge are often falsifiable (Dwyer et al., 2014). In this manner, critical thinking is not only an outward process of finding solutions to problems, but also contains a reflective component that aids further problem-solving.

Creativity

Another component of HOTS that has been greatly discussed in literature pertains to the skills of creativity and innovation. Creativity and innovation have also been discussed within the KSAVE framework similar to critical thinking. Binkley et al. (2012) outlined the requisite knowledge as being able to think and work creatively with others whereupon an individual is aware of the various techniques of idea generation and can utilize them within the constraints of reality, and to implement innovations where the individual understands the impact of innovation and the related boundaries according to history and culture. The related skills are thinking creatively which has to do with generating new, useful ideas, working creatively with others in the development and implementation phases of creativity, and implementing innovations in the sense of developing impactful and feasible innovations (Binkley et al., 2012). Finally, the attitudes and values associated with creativity involve thinking creatively by being open to new ideas, working creatively with others by being open to new and diverse opinions, and implementing innovations by being persistent in presenting new ideas and promoting them to others (Binkley et al, 2012).

Creativity was also a construct measured as part of the HOTS research conducted by Budsankom et al. (2015). In their study, creativity was termed creative thinking and defined as "thinking competency in using previous knowledge to create new knowledge for discovering or innovating new things" (p. 2641, Budsankom et al., 2015) and was linked to outcomes that were valuable for problem-solving. This can relate to Raiyn and Tilchin's (2015) paper on developing HOTS through the use of an adaptive problem-based learning syllabus. In their proposed model, creativity was defined in terms of creative thinking skills which included problem identification, idea generation (termed efficiency), flexibility in producing a wide variety of ideas, originality in terms of producing unusual or uncommon ideas, and elaboration which is seen as the ability to develop ideas fully (Raiyn & Tilchin, 2015). The use of a flexible problem-based learning syllabus with dynamic assessments to suit various HOTS were seen as the way to best increase HOTS development in students (Raiyn & Tilchin, 2015).

In a similar method of using teaching to structure HOTS, Rooney (2012) used inquiry-based learning in an action research to increase HOTS among Mathematics students in Ireland. In this study, creativity was defined using the Anderson and Krathwohl (2001) revised taxonomy and this definition is given as "putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing" (p. 105, Rooney, 2012). Through a sequence of graded assessments such as class presentations and write-up of Mathematics projects, Rooney (2012) found that not only was inquiry-based learning more engaging for students as compared to didactic methods, inquiry-based learning was more beneficial in increasing student HOTS.

Another classroom-based intervention to increase levels of creative thinking was studied by Oncu (2016) where university students were placed in course titled 'Fostering of Creativity and Creative Thinking'. In this study, creative thinking was defined as "a cognitive process of solving problems, generating useful ideas and producing plans that are not present before" (p. 517, Hargrove, 2013, as cited in Oncu, 2016). The Torrance Test of Creative Thinking (TTCT) Figural Form A and B was used to assess the development of creative thinking in students of the course (Oncu, 2016). The TTCT consists of three main activities; the first one requires the respondent to create a drawing with a given shape and the shape must be an integral part of the resulting drawing, the second requires the respondent to complete ten incomplete shapes as a shape or picture, and the last consists of lines or circles over three pages that the respondent needs to use within a picture of their own creation (Oncu, 2016). Post-test results from the study showed that creative thinking of the students increased above national percentages after attending the course, and the improvements were significant particularly in the measures of originality and closure factors (Oncu, 2016).

The TTCT is a commonly used instrument to measure creative thinking that was also used in a study by Moreau and Engeset (2016) to demonstrate how well-defined problem-solving has the capacity to diminish creative and divergent thinking. Creative thinking in their study was defined as a "special class of problem solving characterized by novelty, unconventionality, persistence, and difficulty in problem formulation" (p. 20, Newell, Shaw, & Simon, 1962, as cited in Moreau & Engeset, 2016). In their three-experiment study, they showed that the use of well-defined problems has a carry-over effect in subsequent presentations of ill-defined problems which require creative thought where creativity is inhibited, looking for a correct or 'right' solution also diminishes creativity, and that being in a well-defined-problem-mindset lead to the search for further well-defined problems as the second problem is seen as more feasible and predictable (Moreau & Engeset, 2016).

Another experimental study looked at using online cognitive stimulation tools in the effort to increase creative problem-solving. In this manner, creativity and problem-solving are combined in one construct and creativity is seen as a pathway for problem-solving (Althuizen & Reichel, 2016). In this study, creative problem solving is seen as a cyclical process consisting of six steps; presentation of the problem, preparation for the problem, generation of ideas, evaluation of ideas, selection of 'best' idea, and consequent implementation of the idea (Althuizen & Reichel, 2016).

The phase in which creativity comes most into play would be the idea generation phase as, according to the dual pathway model presented in the paper, individuals require stimulation in order to generate more new ideas by persistently exploring their existing knowledge base or to search for ideas in other categories than the ones they're familiar with (Althuizen & Reichel, 2016). In this manner, creativity in idea generation looks at deepening exploration within a single area of knowledge or broadening exploration across various areas of knowledge (Althuizen & Reichel, 2016). 143 postgraduate students from a business course in Austria were exposed to the use of three types of IT-enabled tools to stimulate creativity, namely a stimuli provider, a process guide, and a mind mapper (Althuizen & Reichel, 2016). They were then asked to provide solutions to a known business problem, the MacCrimmon and Wagner problem of improving a failing donut franchise, where their responses were judged by experts based on the criteria of overall creativity, novelty, and usefulness (Althuizen & Reichel, 2016). It was found that the stimuli provider and the process guide were both useful in increasing the participants' exploration of knowledge thus increasing the total number of ideas generated whereas the mind mapper showed insignificant effects as compared to completely unaided participants (Althuizen & Reichel, 2016). However, overall, the stimuli provider was shown to the most effective in stimulating students to produce a large number of novel and useful ideas through deep exploration of the existing knowledge base (Althuizen & Reichel, 2016).

An earlier study by Carmeli, Gelbard, and Reiter-Palmon (2013) examined similar constructs; the study looked into the relationship between leadership, creative problem-solving capacity, and creative performance. Their definition of creativity integrated definitions by previous researchers whereupon creativity was defined as "the production of new or novel ideas that are useful (Amabile, 1988) and entails change and behavior that defies the norm (Sternberg, 2006)" (p. 97, Carmeli et al., 2013). However, the authors drew a distinction between creativity and creative problem-solving capacity, the latter of which will be discussed in relation to general problem-solving in the subsequent section.

In measuring creative performance, participants were asked to generate as many uses as possible for an object that was familiar to them from the scope of their work (Carmeli et al., 2013). The responses were graded in terms of fluency, as in the total number of ideas generated, and originality, where the idea was extremely uncommon among the total responses (Carmeli et al., 2013). Grading was done by two experts of creativity who were allowed to discuss before consensus was achieved (Carmeli et al., 2013). This assessment was administered to 130 full-time technical employees from an organization that provides utility services (Carmeli et al., 2013). As the study focused on the role of leadership in mediating outcomes of creative problem-solving capacity and creative performance, it was found that levels of internal knowledge sharing between supervisors and subordinates contributed greatly to creative performance (Carmeli et al., 2013).

A recent study examined developmental changes in creativity in 4854 Polish students who were between the ages of four to 21 (Gralewski, Lebuda, Gajda, Jankowska, & Wiśniewska, 2016). In this study, the cross-sectional survey method was used to examine the development of creativity in Polish students from pre-school to university based on the definition of creativity given in Urban (1991, as cited in Gralewski et al., 2016). In Urban's definition, creativity consisted of six criteria:

1. the ability to create a new, unusual and surprising product as a solution to an insightfully perceived problem or a given problem whose implications have been insightfully perceived,
2. and by means of an insightful and broad perception of existing and open data and information purposively looked for,
3. and by analysis, by solution-oriented but highly flexible processing, by unusual associations and new combinations of data and information and with the help of data from experience or with imaginative elements,
4. by synthesizing, structuring and combining these data, elements and structures into a new solution-gestalt (whereby the processes in 3. and 4. may partially run simultaneously on different processing and consciousness levels),
5. to arrive at a solution-gestalt, which as a product or in a product, in whichever form, becomes elaborated,
6. and finally through communication can be grasped via the senses and experienced by others as meaningful and significant (p. 104-105, Urban, 1991).

In this definition, it can be seen that several elements of HOTS are integrated in the construct of creativity such as analysis, synthesis, and problem-solving. Based on this, the Urban-Jellen Test of Creative Thinking – Drawing Production (TCT-DP) was used as a measure of creativity in Gralewski et al.'s (2016)

study. In the test, respondents were required to form a drawing out of six objects placed asymmetrically on the drawing space (Gralewski et al., 2016). The final drawing is assessed on several criteria:

> Continuations (Cn), Completions (Cm), New elements (Ne), Connections made with a line (Cl), Connections that con-tribute to a theme (Cth), Boundary breaking that is fragment-dependent (Bfd), Boundary breaking that is fragment-independent (Bfi), Perspective (Pe), Humour and affectivity (Hu), Unconventionality with subcriteria (Uc) [(a) manipulation of the test material (Uca); (b) surrealistic or abstract elements (Ucb); (c) use of symbols or signs (Ucc); (d) unconventional usage of the given fragments (Ucd)] and Speed (Sp) (p. 158, Gralewski et al., 2016)

The score of all criteria is summed to produce the final score (Gralewski et al., 2016). From the study, it was found that 13 of the 14 criteria showed changes across age levels, with only Unconventional manipulation of test material (Uca) showing no change (Gralewski et al., 2016). Overall, a non-linear pattern of development was seen across the various criteria, with increases and decreases occurring non-uniformly among the criteria (Gralewski et al., 2016). However, across all 14 criteria and the total TCT-DP score, a slump in creativity was seen in the adolescent phase beginning in middle school and continuing throughout high school before picking up again in university (Gralewski et al., 2016).

From the aforementioned review, it can be seen that creativity has been assessed in a multitude of ways. Though many tests of creativity and creative thinking have been developed, a study by Runco, Abdulla, Paek, Al-Jasim, and Alsuwaidi (2016) sought to examine which could be considered the best measurement. In this study, seven measures of divergent thinking were identified and administered to 611 participants from three universities within the Gulf region in a one-shot post-test only design (Runco et al., 2016). Divergent thinking is seen as a form of measurement of creative potential, and is operationalised as the "number, originality, and flexibility of ideas produced" (p. 5, Runco et al., 2016). Of the seven tests administered, four were sourced from the Runco Creative Assessment Battery, specifically the Figural Divergent Thinking, Titles, Realistic Presented Problems, and the Realistic Problem Generation tests, whereas the other three tests came from the Wallach and Kogan battery, specifically the Instances, Uses, and Similarities tests (Runco et al., 2016).

In the Figural Divergent Thinking test, participants are presented with a figure and asked to list down as many things as possible that they think it represents (Runco et al., 2016). In the Titles test, they're presented with the paragraph of a story and asked to generate as many alternative titles as they can (Runco et al., 2016). For the Realistic Presented Problems, they are presented with a description of a problem that may happen at school or work and they're required to list down as many possible solutions as they can, and in the Realistic Problem Generation, they are given an open-ended question about general life situations and asked to list as many problems as they can that they think could be encountered (Runco et al., 2016). For the Wallach and Kogan Instances test, participants had to list out all the items they could think of that fit a given criterion (Runco et al., 2016). In the Similarities test, they were requested to list out all the similarities between a pair of given objects, and in the Uses test, they were tasked to list out all the uses possible for certain common household objects (Runco et al., 2016). All responses were analysed according to the degree of originality, where the authors determined the cut off for original ideas as those given by approximately 5% of the sample (Runco et al., 2016).

Aside from the tests of divergent thinking, participants were also administered the Quick Estimate of Convergent Thinking (QECT) as a kind of check to ensure the divergent thinking tests were really measuring divergent thinking (Runco et al., 2016). A low or negative correlation between scores of the divergent thinking tests and the QECT would be indicative that the divergent thinking tests are operating as they should (Runco et al., 2016). In addition, the Runco Ideational Behaviour Scale (RIBS) was also administered to capture the participants' self-reports of ideational behaviour as it is based on the belief that ideas are products of creative thinking (Runco et al., 2016). A five-point Likert scale was used to provide ratings of how often the participants engaged in behaviour that was related to having ideas (Runco et al., 2016).

Based on the results of the test administration, the researchers found that the Figural Divergent Thinking and Wallach and Kogan's Uses tests showed the lowest levels of reliability (Runco et al., 2016). The Titles and Realistic Problem Generation tests demonstrated the highest levels of originality across all tests but overall, the researchers recommended that more than one test be used to assess creativity as not all tests were equally generalizable (Runco et al., 2016).

Problem-solving

Problem-solving often shows overlap with the other constructs of creativity (as in creative problem-solving) or as a by-product of critical thinking. In the study by Moreau and Engeset (2016), problem-solving and creativity were linked in that solving well-defined problems had a negative impact on creativity. Problem-

solving here was defined as occurring within a problem space which is a mental representation or the structure of a problem within the mind of the solver (Moreau & Engeset, 2016). The Miller Analogy Task was used to represent a well-defined problem in this study. Where the Torrance Test of Creative Thinking presented an ill-defined problem (and an opportunity for creative thought), the Miller Analogy Task presents 25 analogy questions in which participants need to answer using reasoning by analogy (Moreau & Engeset, 2016). Though the authors note that it is a well-defined problem which is often used to test problem-solving skills, there is ambiguity (and, hence, the opportunity for creativity as well) in which the participants can use their own forms of analogical thinking to reach a solution (Moreau & Engeset, 2016).

Similarly, in the 2013 study by Carmeli et al., the authors drew a distinction between creative problem-solving capacity and creative performance. Creative problem-solving capacity differed from typical definitions of creativity or creative performance in that consisted of the basic cognitive processes that are often associated with idea generation "that includes identification and construction, information search and acquisition, and ideation, as well as the implementation phase, which includes idea evaluation, idea selection, and implementation planning" (Reiter-Palmon & Illies, 2004, as cited in p. 97, Carmeli et al., 2016). In this, it can be seen that creative problem-solving capacity contains elements of analysis and evaluation in addition to problem-solving and creativity.

To assess the degree of creative problem-solving capacity in the participants, an author-made scale of eight items with every two items measuring a different process was constructed (Carmeli et al., 2016). The four processes measured in the scale included "problem construction and identification, idea generation, idea evaluation, and idea implementation" (p. 101, Carmeli et al., 2016). A five-point Likert scale was applied where respondents indicated the degree to which they possessed the capabilities represented in the items (Carmeli et al., 2016). Their study found that information sharing within the organization and between organization members and external parties improved creative problem-solving capacity, and that creative problem-solving capacity was a mediator of the relationship between leadership and creative performance as discussed in the preceding section (Carmeli et al., 2016).

Beyond traditional problem-solving, a study by Hwang and Kuo (2015) examined web-based problem-solving performance, a skill which is becoming more in demand given the need for information technology proficiency in the modern world. Using structural equation modelling, they sought to uncover the antecedents to web-based problem-solving to better understand the factors that contribute to this skill (Hwang & Kuo, 2015). Hwang and Kuo (2015) defined web-based problem-solving in their study as a process of higher-order thinking. It was seen as a combination of critical thinking, creativity, cognitive reasoning, and web-based information search skills (Hwang & Kuo, 2015). They proposed a five-phase learning cycle consisting of constructing prior knowledge which refers to the existing contextual information surrounding the problem, adopting keywords which are terms relevant to defining the problem, identifying information which is the search for relevant information and elimination of useless solutions, information abstraction which refers to extracting information that can be used for problem-solving, and elaborating on thinking which is to present ideas and reflect upon their usefulness in order to refine the final solution (Hwang & Kuo, 2015).

In assessing problem-solving, the authors used a 10-item measure that evaluated the students' problem-solving ability along the following aspects: "awareness of problems, identification of the nature of problems, recognition of factors related to the problems, identification of more information needed to solve the problems, and determination of solutions" (p. 407, Hwang & Kuo, 2015). An overall score of 20 could be achieved and the higher the score, it was assumed the better the problem-solving ability of the student (Hwang & Kuo, 2015). The authors also ensured that inter-rater reliability was maintained by enlisting four primary school level social science teachers from different institutions to rate the students (Hwang & Kuo, 2015). Utilizing a pre-test-post-test experimental design, Hwang and Kuo (2015) assessed 201 primary school students between the ages of 11 and 12 from three elementary schools in Southern Taiwan who were exposed to a web-based learning environment. Not only did they find empirical evidence of construct validity for their proposed model of web-based problem-solving performance through the use of structural equation modelling, they found that the main factor affecting students' intention to learn via the Internet (which subsequently had a direct effect on problem-solving performance) was task-technology fit, which refers to goodness-of-fit between the available technology and its use in completing a given task (Hwang & Kuo, 2015).

Most theoretical frameworks that describe problem-solving often define it in the manner of a process. For example, a positional paper by Greiff et al. (2014) which discusses methods to increase domain-general problem-solving skills through the use of education examines problem-solving as the cognitive capacity to engage in, understand, and find solutions to a situation where one might not be obvious. Beyond the act of mere problem-solving, it encompasses an individual's intention to engage with a problem before moving on to other processes such as being able to understand the entire situation and various aspects of the problem, identifying the main problem and sub-problems that need to be solved, hypothesizing about methods that can be used as

solutions, planning and executing a solution of choice before monitoring and evaluating outcomes as they unfold (Grieff et al., 2014). The authors specifically discussed the use of PISA in assessing general problem-solving and presented a proposed framework to leverage on education in improving these skills as they believed researchers had a responsibility to increase awareness among relevant stakeholders, develop optimized assessments to measure these skills, and explore new methods of developing these skills among students (Grieff et al., 2014).

A theoretical framework that has been tested within the Malaysian context is the Marzano HOTS framework. In a study by Yee et al. (2011), the researchers sought to determine the level of HOTS according to the Marzano framework in Technical and Vocational students at a local public university. Within the Marzano framework, problem-solving is defined as "overcoming constraints or limiting conditions that are in the way of pursuing goals" (p. 122, Yee et al., 2011). The Rubrics for Specific Task or Situation, which is a 44-item scale based on the 13 Marzano HOTS, was distributed to 158 students of the Technical and Vocational Education course at a local public university (Yee et al., 2011).

It was found that students perceived they were at moderate levels for seven out of 13 HOTS (investigation, experimental inquiry, comparing, deducing, constructing support, inducing, and invention) but were at low levels for the remaining HOTS (decision making, problem solving, error analyzing, abstracting, analyzing perspectives, and classifying); in addition, no statistically significant differences were found across gender, academic achievement, and socioeconomic status (Yee et al., 2011). In this manner, it was concluded by researchers that students needed more assistance in developing their HOTS either in formal teaching and learning environments, or through self-study (Yee et al., 2011).

Analysis

Within the three areas of HOTS discussed in the earlier sections, a few major skills have been seen to be consistently appearing as integral to the overall thought process. The first to be discussed would be analysis. Most recently, a paper by Nor'ain and Chinnappan (2016) reviewing the theoretical links between HOTS and performance in TIMSS (Trends in International Mathematics and Science Study) proposed that HOTS were crucial for students in order to properly transfer their learning onto novel, unusual problems such as those presented in TIMSS. In this paper, they defined analysis as the ability to determine the relevant elements in a problem as well as the links between them (Nor'ain & Chinnappan, 2016). Overall, they analysed several TIMSS tasks to provide Mathematics teachers with an understanding of how HOTS can enable students to move to higher cognitive levels within the context of TIMSS tasks (Nor'ain & Chinnapan, 2016).

Elements of this definition of analysis are shared in Budsankom et al.'s (2015) definition of analytical thinking which is given as the ability "to classify objects logically, assessing the relationships of certain elements, how they contribute, how they relate to each other, how they work, and what the most important parts are" (p. 2641). Another theoretical paper in the same year by Raiyn and Tilchin (2015) present a differing view of analysis. Here, analytical thought or logical thought is seen as a group of skills which allow for critical thinking to occur, and are essential for choosing the optimum solution to a problem (Raiyn & Tilchin, 2015). The skills encompass "ordering, comparing, contrasting, evaluating and selecting" (p. 93, Raiyn & Tilchin, 2015).

Dwyer et al.'s (2014) theoretical framework includes analysis as a skill within the realm of critical thinking. Analysis is defined as a skill related to argumentation, whereupon it is used to identify and scrutinize the propositions of an argument, and their roles in the overall strength of the argument (Dwyer et al., 2014). Analysis is also related to being able to understand the structure of arguments, taking into consideration the contextual and linguistic cues aside from the pure semantic content (Dwyer et al., 2014). In this manner, analysis is process of dissecting an argument but in a holistic manner that also accounts for the gist of what is being said.

In assessing analysis, a 2013 study by Momsen et al. examined the differences in cognitive levels tested by introductory physics and biology examinations for students in an undergraduate course in America. Though their study was more related to tailoring classroom instruction in accordance to the levels of Bloom's taxonomy used in assessments, they provided an interesting perspective in using regular classroom assessment to measure HOTS (Momsen et al., 2013). Analysis by their definition was related to being able to pick out patterns, recognize underlying or hidden meaning, and detecting and organizing components within a system (Momsen et al., 2013). However, in the courses they sampled, knowledge, comprehension, and application took precedence over analysis in student assessments (Momsen et al., 2013).

The idea of deconstructing systems that was presented in Momsen et al's (2013) definition was mirrored in the definition of Ramos et al. (2013) who also based their HOTS assessment on usual Physics classroom tests. In this study, analysis was multifaceted; beyond the ability to understand relationships between

individual parts and the whole, and the ability to comprehend causal relationships, analysis involved being able to glean information from pictorial representations of data and reflectively restructuring knowledge in novel ways (Ramos et al., 2013). Their definition seemed to be both reductionist and holistic at once, where the individual is able to not only identify individual parts but understand how everything fit together in order to work (Ramos et al, 2013).

Kim, Patel, Uchizono, and Beck (2012) presented a study that sought to incorporate elements of Bloom's taxonomy into a pharmaceutical care classroom. Higher-order elements of Bloom's taxonomy such as synthesis, evaluation, and creation were added to the syllabus through the use of multiple-choice examination questions that were used to determine the students' letter grades for the compulsory course on therapeutics (Kim et al., 2012). In this study, analysis was defined as "the ability to break down the materials into its constituent parts and detect the relationships of the parts and of the way they are organized" (p. 3, Kim et al., 2012).

When scoring the assessments, synthesis and evaluation questions had to be combined as the authors noted that the hierarchical order of these two elements were often reversed in literature; for example, Anderson and colleagues placed evaluation before creation where in the original, synthesis preceded evaluation (Kim et al., 2012). Their study found that questions with elements of application and synthesis were better at differentiating high performance and low performance students compared to just knowledge and comprehension questions (Kim et al., 2012). Thus, it was able to improve the discrimination value of the test without affecting the average scores in addition to being slightly more difficult with the various aspects tested (Kim et al., 2012).

Rooney (2012) also based her study on Anderson and Krathwohl's (2001) revision of Bloom's taxonomy. Her action research was guided by their definitions of each skill; analysis was defined as "breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing" (p. 105, Rooney, 2012). Pardamean (2012), however, takes a more holistic approach in defining analysis such that he emphasizes the comprehension and expression of the meaning behind experiences, beliefs or situations, and this comes in hand with identifying the manifest and latent or implied relationships between the components of representations of said experiences and beliefs. This view differs greatly from the rest as it is the only definition that does not seek to 'break down' or reduce the object down to the constituent parts.

Evaluation

The next major skill would be evaluation. This skill has been seen in most of the studies previously discussed. In defining evaluation with relation to Mathematics tasks of the TIMSS, Nor'ain and Chinnappan (2016) suggested that evaluation is the process of examining the 'reasonableness of values x and y' (p. 205). The Dwyer et al. (2014) framework looks at evaluation as a form of assessment. As the framework examines thinking skills in relation to proposed arguments, evaluation is about judging the points and inferred conclusions of an argument based on the strength of its logic, its relevance to the point, overall credibility, and potential for bias or omissions, which requires holistic examination of not just the argument itself, but the delivery and trustworthiness of the deliverer (Dwyer et al., 2014).

Momsen et al. (2013) offer a more succinct definition that's relevant to scientific inquiry. Evaluation here is defined as the ability to compare and discriminate between ideas, to assess the worth of theories and hypotheses, to make decisions with respect to logical and reasonable arguments, and to confirm the credibility of evidence and understand that subjectivity exists in all arguments (Momsen et al., 2013). With the Ramos et al. (2013) definition, evaluation also incorporates elements of expression and defence of arguments in addition to the process of judging quality and strength using a predetermined criteria. Kim et al. (2012) base their definition from that of Bloom's taxonomy where evaluation is a judgmental process of assessing the worth of ideas, work output, or solutions.

Rooney (2012) uses the Anderson and Krathwohl (2001) revision of Bloom's taxonomy which defines evaluation as "making judgments based on criteria and standards through checking and critiquing" (p. 105). Pardamean (2012) shares a similar definition to Dwyer et al. and Momsen et al. but adds that evaluation requires the individual to provide the outcomes of the reasoning process and justify this process in regard to evidence, concepts, methods, and other contextual considerations.

Synthesis

Synthesis is a skill that is often combined with other skills, or simply regarded as a part of creation (creativity). As such, few studies provide full definitions of synthesis. However, it can be seen in studies that base their assessments on Bloom's taxonomy. This is often reflected more in classroom or subject-specific assessments. Momsen et al. (2013) define synthesis as the ability to create new ideas on the basis of older ones, to generalize from a set of given facts, and to understand and use relationships of knowledge from different

subject areas. Kim et al. (2012) define synthesis as "the ability to put parts together to form a whole, with emphasis on creating a new meaning or structure" (p. 3).

Conceptual Framework

From the review of literature, it was seen that HOTS are often measured or seen as three major areas; critical thinking, creativity, and problem-solving. However, there is an overlap in the three areas, particularly with regard to creativity and problem-solving. In assessing creativity, most assessments either measure creative potential/capacity or creative performance. The overlap is seen with problem-solving because it requires creativity to formulate new solutions to a given problem. The other skills of analysis, evaluation, and synthesis are mostly embedded within the major three areas and are, thus, often assessed simultaneously. Thus, the current toolkit placed a focus on assessing these three umbrella concepts as being representative of HOTS. Figure 1 depicts the conceptual framework of this study.

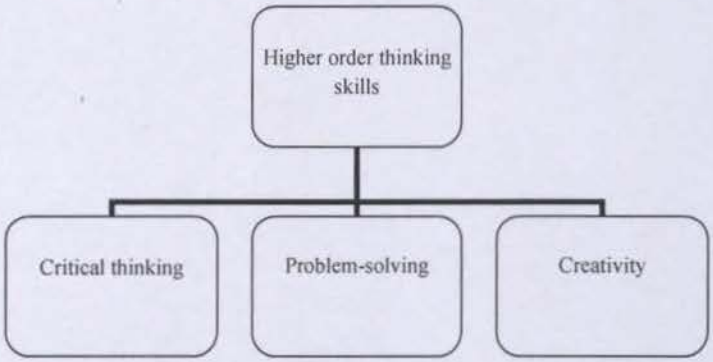

Figure 1: Conceptual Framework of HOTS Assessment Toolkit.

METHOD

This section will present the framework of test development used in this study, as well as describe the current progress made. Following that, a brief description of future directions will be provided.

Scale Development

In this study, the focus lay on the development of the critical thinking and problem-solving scales. The creativity scale was adapted from a pre-existing instrument and its adaptation will be discussed in a separate paper. The development of the critical thinking and problem-solving scales was carried out in accordance with DeVellis' (2003) guidelines of scale development. DeVellis provided an eight-step framework of test development, as illustrated in Figure 2.

In the first stage, the constructs to be measured were decided on by conducting a review of relevant literature. Operational definitions for each construct were developed based on a synthesis of the definitions of extant literature. Within each construct, certain skills were identified as necessary in each process. Table 1 presents the operational definitions of each construct with the requisite skills.

Based on these operational definitions, test items were generated for critical thinking and problem-solving. The format of measurement was determined prior to writing items as the type of items used informed the writing process. Given the non-cognitive nature and non-specificity of its measurement in terms of content area in the curriculum of the assessment toolkit, syllogisms and situational judgement test (SJT) items were generated.

The Critical Thinking scale comprises of syllogisms which require the test-taker to select the best conclusion to an argument based on the strength of their logical reasoning and evaluation of a set of given premises. This scale consists of 30 items which are presented in the multiple-choice format with four response options. Following Peirce's theory of syllogisms (1867), deductive argumentation has been found to be a good predictor for college success.

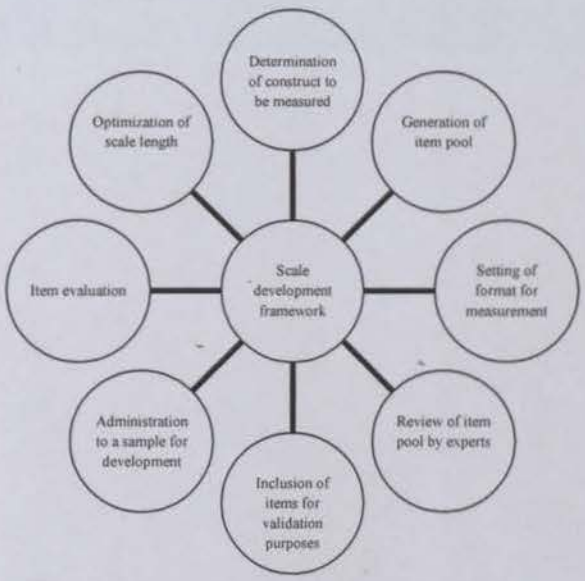

Figure 2: DeVellis' (2003) Guidelines for Scale Development.

Table I: Operational Definitions for Critical Thinking, Problem-solving, and Creativity.

Critical thinking is an evaluative, logic-based cognitive process with the aim of supporting a proposed argument based on gathered information	Problem-solving is a cognitive decision-making process where the skills of critical thinking are used to determine the optimal solution to a presented problem taking into account possible existing or upcoming barriers at all stages of solving the problem	Creativity is defined as a two-part process where creative thinking is the precursor to creative production but can occur independently whereas creative production is entirely dependent on creative thinking
Skills involved:	Skills involved:	Skills involved:
Smart searching for new information based on available information which involves using keywords to obtain new, relevant information	Smart searching for new information based on available information which involves using keywords to obtain new, relevant information	Creative thinking is the generation of novel ideas. It involves the skill of synthesis where existing, available information is combined in novel and unique permutations, or new information is produced and added to existing information to produce unique or uncommon ideas. Breadth of creative thinking is estimated by fluency (number of ideas generated) whereas depth of creativity is estimated by originality (uncommonness of ideas generated) and flexibility (spread of differentiation between ideas generated). At this stage, no evaluation is done on the practical worth or value of the ideas generated; all ideas are equivalently important and none are discarded
Analysis of obtained information by examining each unit of information, and discerning relationships as well as the underlying meaning of relationships between units of information	Analysis of obtained information by examining each unit of information, and discerning relationships as well as the underlying meaning of relationships between units of information	
Evaluation of units of information by logically considering the value of information in relation to its credibility, feasibility, and usefulness	Evaluation of units of information by logically considering the value of information in relation to its credibility, feasibility, and usefulness	
Synthesizing new networks or permutations of information based on logical transformation of available information	Synthesizing new networks or permutations of information based on logical transformation of available information	
Generalizing available information beyond the boundaries of a specific context to other contexts	Generalizing available information beyond the boundaries of a specific context to other contexts	Creative production involves logically evaluating the feasibility and practical worth (usefulness) of ideas generated and expanding them before bringing them into reality. Each idea is judged on its
Using logical reasoning to apply informational output from the thought process in order to provide	Generating possible solutions to the presented problem using synthesized information as the	

reasonable explanations in support of a proposed argument	stimulus to idea generation, ideas may be novel and unique or adaptations of known solutions	depth, usefulness, and real-world feasibility. Ideas of higher worth are systematically expanded through logical analysis of strengths and weaknesses until they can be produced within the constraints of reality. Creative production requires reflective thinking where unique, novel ideas are developed within the constraints of reality while working around personal limitations to resources, techniques, and skills.
	Using logical reasoning to apply informational output from the thought process in order to provide potentially effective solutions to the presenting problem or discard potentially ineffective solutions	

The Problem-solving scale, on the hand, comprises of Situational Judgment Test (SJT) items, where test-takers choose the best response to a certain situation based on their reasoning skills. There are 10 situations, with five pertinent items/questions each, in a 4-option multiple-choice-question format. Situational Judgment Tests (SJTs) have been found to be reliable assessments in recruitment and personnel selection as well as college admission tests over the past two decades. This type of test had been found more valid and beneficial compared to other simulation-based assessments. Recent research has demonstrated that SJTs possess a number of positive features, such as predictive and incremental validity, small racial and gender subgroup differences and favorable user reactions (Kasten & Freund, 2016).

In both scales, participants are required to choose the best answer from a set of four options. In the Critical Thinking scale, the best answer would be the one that follows logically from the given premises. In the Problem-solving scale, the best answer is decided upon based on consensus from the expert review. At this current stage, the items have been sent out for expert review. This crucial step will establish the content validity and readability of the items/questions, given the ability of the target population. The inputs of the experts will be considered to improve the test items/questions, as well as clarifying whether or not the tasks asked in the tests are clear and plausible enough with the given options per item/question.

FUTURE DIRECTIONS

The study will undergo a series of pilot-testing and several iterations of item analysis using the Rasch model. The item banking must be large enough that when an item replacement is needed, there are alternative items/questions ready for use. Validation studies as to its concurrent and predictive validity will be another phase of this study. Once desirable psychometric properties are attained, the assessment toolkit must be rolled out to its target population, and further studied for its reliability and stability over time.

ACKNOWLEDGEMENT

This paper is part of a project to develop higher order thinking skills toolkit for secondary school learners. The project is funded by the Fundamental Research Grant Scheme (FRGS) under the Ministry of Education, Malaysia, with Grant Number: FP029-2016.

REFERENCES

Althuizen, N. & Reichel, A. (2016). The effects of IT-enabled cognitive stimulation tools on creative problem solving: A dual pathway to creativity. *Journal of Management Information Systems*, *33*(1), 11–44.

Azura Abas. (2016, 16 December). Heated up over 'HOTS skills'. *New Straits Times*. Retrieved from https://www.nst.com.my/news/2016/12/197229/heated-over-hots-skills.

Barnett, J. E. & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, *32*(2), 201-211.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century skills. In P. Griffin, B. McGaw, &E. Care (Eds.), *Assessment and teaching of 21st Century skills* (pp. 17-66). United States of America: Springer. doi: 10.1007/978-94-007-2324-5.

Budsankom, P., Sawangboon, T., Damrongpanit, S., & Chuensirimongkol, J. (2015). Factors affecting higher order thinking skills of students: A meta-analytic structural equation modeling study. *Educational Research and Reviews*, *10*(19), 2639-2652.

Butler, H. (2012). Halpern Critical Thinking Assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, *26*(5), 721-729.

Carmeli, A., Gelbard, R., & Reiter-Palmon, R. (2013). Leadership, creative problem-solving capacity, and creative performance: The importance of knowledge sharing. *Human Resource Management*, *52*(1), 95-122.

Çavdar, G. & Doe, S. (2012). Learning through writing: Teaching critical thinking skills in writing assignments. *Political Science and Politics*, *45*(2), 298-306.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd Ed.). California: Sage Publications, Inc.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52.

Gralewski, J., Lebuda, I., Gajda, A., Jankowska, D. M., & Wiśniewska, E. (2016). Slumps and jumps: Another look at developmental changes in creative abilities. *Creativity: Theories-Research-Applications*, *3*(1), 152-177.

Greiff, S., Wüstenberg, S., Csapó, B., Demetriou, A., Hautamäki, J., Graesser, A. C., & Martin, R. (2014). Domain-general problem solving skills and education in the 21st century. *Educational Research Review*, *13*, 74-83.

Hwang, G-J., & Kuo, F-R. (2015). A structural equation model to analyse the antecedents to students' web-based problem-solving performance. *Australasian Journal of Educational Technology*, *31*(4), 400-420.

Kasten, N. & Freund, P.A. (2016). A meta-analytical multilevel reliability generalization of Situational Judgment Tests (SJTs). *European Journal of Psychological Assessment*, *32*(3). DOI:10.1027/1015-5759/a000250.

Kim, M-K., Patel, R. A., Uchizono, J. A., & Beck, L. (2012). Incorporation of Bloom's taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. *American Journal of Pharmaceutical Education*, *76*(6), Article 114, 1-8.

Ministry of Education. (2013). *Malaysia Education Blueprint 2013-2025 (Preschool to Post-Secondary Education)*. Putrajaya: Ministry of Education Malaysia.

Momsen, J., Offerdahl, E., Kryjevskaia, M., Montplaisir, L., Anderson, E, & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate Science courses. *CBE – Life Sciences Education*, *12*, 239-249.

Moreau, C. P. & Engeset, M. G. (2016). The downstream consequences of problem-solving mindsets: How playing with LEGO influences creativity. *Journal of Marketing Research*, *LIII*, 18–30.

Naj. (2017, 30 March). Lower Number of Straight As For SPM 2016, We Ask Experts If The New HOTS System Works. *Malaysian Digest*. Retrieved from http://www.malaysiandigest.com/frontpage/282-main-tile/666425-lower-number-of-straight-as-for-spm-2016-we-ask-experts-if-the-new-hots-system-works.html.

Nor'ain Tajudin, & Chinnappan, M. (2016). The link between higher order thinking skills, representation and concepts in enhancing TIMSS tasks. *International Journal of Instruction*, *9*(2), 199-214.

Oncu, E. C. (2016). Improved creative thinkers in a class: A model of activity based tasks for improving university students' creative thinking abilities. *Educational Research and Reviews*, *11*(8), 517-522.

Pardamean, B. (2012). Measuring change in critical thinking skills of dental students educated in a PBL curriculum. *Journal of Dental Education*, *76*(4), 443-453.

Raiyn, J., & Tilchin, O. (2015). Higher-order thinking development through Adaptive Problem-based Learning. *Journal of Education and Training Studies*, *3*(4), 93-100.

Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher order thinking skills and academic performance in Physics of college students: A regression analysis. *International Journal of Innovative Interdisciplinary Research*, *4*, 48-60.

Rooney, C. (2012). How am I using inquiry-based learning to improve my practice and to encourage higher order thinking among my students of mathematics? *Educational Journal of Living Theories*, *5*(2), 99-127.

Runco, M. A., Abdulla, A. M., Paek, S. H., Al-Jasim, F. A., & Alsuwaidi, H. N. (2016). Which test of divergent thinking is best? *Creativity: Theories-Research-Applications*, *3*(1), 4-18.

Yee, M. H., Widad Othman, Jailani Md Yunos, Tee, T. K., Razali Hassan, & Mimi Mohaffyza Mohamad. (2011). The level of Marzano Higher Order Thinking Skills among Technical Education students. *International Journal of Social Science and Humanity*, *1*(2), 121-125.