

## Contents

Purpose .....	2
Scope.....	2
What is HOT? .....	2
How do we measure HOT? How are they similar and different? .....	3
Bloom’s Taxonomy.....	3
Ennis Critical Thinking .....	4
Solo Taxonomy.....	5
Garrison’s Community of Inquiry.....	5
How have they been applied in settings comparable to what I will be studying? .....	6
Model Choice Considerations .....	7
Units of Analysis.....	11
Exogenous factors affecting HOT demonstration.....	13
Choice of Framework (trying to get at construct validity).....	15
Case study of application to experience as contributor to AnkiDroid.....	16
Case study of application to OpenMRS mailing list .....	16

## **Purpose**

The purpose of this literature review is to identify how higher order thinking (HOT) skills are defined and measured so that for my proposal it can inform the choice of HOT skills I will study and how they can be measured.

## **Scope**

**(might be able to just cover in the application section)**

## **What is higher-order thinking (HOT) skills?**

HOT skills are difficult to define (Lewis and Smith, Resnick). Lewis and Smith (1993) set out to define and differentiate it from other commonly used terms, critical thinking and problem solving. The authors discuss the humanities' focus on critical thinking in the evaluative sense and science and mathematics focus on problem solving, noting that while each domain may focus on certain skills both are needed in all disciplines.

King et. al. in a thorough review of the major learning theorists' concepts of higher-order thinking also notes the importance of researcher background (objectives?) in influencing the definition and effecting confusion. In the end, however, they cite Lewis and Smith's definition and expression of value as the best:

Higher order thinking occurs when a person takes new information and information stored in memory and interrelates and/or rearranges and extends this information to achieve a purpose or find possible answers in perplexing situations. A variety of purposes can be achieved through higher order thinking as defined above. These would include: deciding what to believe; deciding what to do;

creating a new idea, a new object, or an artistic expression; making a prediction; and solving a nonroutine problem.

After examining definitions and various taxonomies they then enumerate the situations and skills involved and outcomes of higher order thinking. Some situations leading to development of HOT include ambiguities, challenges, confusions, and problems involving skills such as complex analysis, reflection, systems analysis, and synthesis. Example outcomes include arguments, confirmations, hypotheses, plans, and solutions.

Resnick (1987) also defines higher order thinking similarly as Lewis and provides more detail about what it looks like, noting that "... although we cannot define it exactly, we can recognize higher order thinking when it occurs." Notably she describes it as being "nonalgorithmic" often having "multiple solutions" that require "nuanced judgment and interpretation" and often involves "uncertainty." The process is one of "imposing meaning", which is an "effortful" process requiring "self-regulation." Finding one all-encompassing definition may not be reasonable and is likely not the place to start when assessing HOT skills, models are the level be the level at which skills should be defined. (McLean) The next section is an overview of a few well known models including their intent when created and general structure.

## **How do we measure HOT?**

### **Bloom's Taxonomy**

While the authors discussed previously tend to refer to HOT in similar ways, there are frameworks that have been developed that can be used as more detailed jumping off points for measuring the exhibition of HOT skills. In general, frameworks vary in their nature of higher order skill measurement and intent for use. Bloom's taxonomy which took eight years to develop, first appearing in 1956, originated from a desire to share test items between faculty at

different institutions and developed into a classification of educational goals and objectives. (Krathwohl, Forehand) The taxonomy covers three domains: cognitive, affective, and psychomotor; for the purpose of this literature review the focus here will be on the cognitive domain. The framework was revised in 2001 to a two-dimensional scale containing knowledge and cognitive process. Behaviors can now be classified according to the knowledge required: factual, conceptual, procedural, and metacognitive and the cognitive process used: remember, understand, apply, analyze, evaluate, and create. The cognitive process is defined as hierarchical with higher levels subsuming lower levels implying more complex processing requires mastery of less complex processes. However, there is some evidence that higher order thinking is not necessarily hierarchical (Resnick (1987), Newman (1990), Herrington (1999), Ennis (1993)).

### **Ennis Critical Thinking**

Ennis does not use the term higher-order thinking, but uses critical thinking instead, making the tie between higher-order thinking and critical thinking by referring to some peoples' definition of critical thinking as being composed of the top three levels of Bloom's taxonomy—analyze, evaluate, create. His purpose for defining critical thinking is for developing proper assessment, which he says is not facilitated by the vaguely defined categories in Bloom's taxonomy. (Ennis)

Ennis' model contains 4 categories of critical thinking – clarity, basis, inference, and interaction.

Clarity relates to indicators such as asking and answering questions for clarification and identifying assumptions. Basis pertains to evaluating evidence and providing support for reasoning. Inference includes induction and deduction as well as the evaluation thereof and “making value judgements.” Interaction means working with others to identify action(s) and evaluating outcomes. (McLean)

## **Solo Taxonomy**

The SOLO model, developed in 1982, is based on Piaget's definition of the cognitive stages of development. It consists of 5 different stages of increasing structural complexity: prestructural, unistructural, multistructural, relational, and extended abstract. Prestructural is the lowest stage defined by student responses indicating a non-understanding of the problem. Unistructural is characterized as having a surface understanding with "one relevant aspect" of the problem realized by the student. In multistructural more than one substantive aspect of the task is understood, however no relationship is formed; this is still considered a surface approach. Multistructural is likened to Bloom's procedural (apply?) level. When students can relate the pertinent information in the task into a whole the relational level is reached, and when that whole is expressed at as a higher abstraction the extended abstract is attained. (Chan)

## **Garrison's Community of Inquiry**

Garrison's Community of Inquiry model arose from a need to help structure education experiences appropriately as higher education began to use computer-mediated communication (CMC) for distributed learning purposes. (Garrison 01) The model consists of three components that are necessary for successful learning when education is conducted in a CMC environment: social presence, cognitive presence, and teaching presence. Social presence relates to the ability for the learner to develop an identity within the community. Cognitive presence is the ability of the community members to be able to construct meaning through the CMC environment. Lastly, teaching presence consists of instructional design and facilitation of the learning process and supports the cognitive and social aspects.

Since this review is concerned with higher-order thinking, only the cognitive presence dimension will be discussed in detail. The cognitive presence element is called the Practical Inquiry (PI)

model in Garrison's framework, named as such because it is based on Dewey's concept of practical inquiry, which relies heavily on reflection. Because of a constructivist approach to the educational process the model is general is described as having an "iterative and reciprocal relationship between personal and shared worlds. ... This reflects the dynamic relationship between personal meaning and shared understanding (i.e., knowledge)". (Garrison 01)

As with the Bloom and SOLO taxonomies it consists of multiple phases. The process begins with a triggering event which generates a state of unease such as a problem, confusion, or question. Exploration is the second category and may involve looking for more information about and/or discussing the details of the triggering event. Integration, the third category, is about digesting(hmmm, wording) the information to formulate an overall idea or concept. Lastly, resolution is an attempt to apply ideas to resolve the issue and critically evaluate solutions.

### **How have HOT frameworks been applied?**

While free and open source software (FOSS) communities have been studied on many fronts including member characteristics, project characteristics, social processes, software development processes and FOSS use (12 Crowston), there appears to be no research concerning exercise of higher-order thinking as a participant in FOSS. Such information would point to the potential for students to exercise those skills through participation in OSS. As such, this became one of the research questions in my prospectus. So, to inform the methodology for study of HOT in open source software communities my literature search turned to studies examining those skills in domains with similar characteristics.

This research review is primarily related to study of HOT within online educational environments and application of HOT frameworks to defining software engineering skills. HOT skills in online education was chosen because the area I plan on eventually studying, OSS, has some characteristics in common with online educational environments; OSS uses threaded discussions as does the studies reviewed and both are socio-constructivist learning environments. (cite, are the studies I reviewed looking at socio-constructivism) Furthermore, the studies discussed here all pertain to undergraduate and graduate students so it is applicable to the learner of interest. Skill identification in software engineering using cognitive frameworks was also reviewed because work in that domain is similar to OSS.

All the studies related to online education used content analysis as the major investigation mechanism. Themes that emerged related to conducting the content analysis were: 1) selection of the right model(s) for the domain and measuring quality, 2) unit of analysis; and 3) exogenous factors that influence exercise and measurement.

### **Model Choice Considerations**

Note: maybe add in here about triangulation and recommendations by authors, or perhaps note in my methodology

04 garrison

Fahy?

Meyer

Herrington 99 in studying the ability to exercise HOT in a class using multimedia for education of pre-service math teachers first began with enumerating the characteristics of HOT as defined by Resnick, then corroborated HOT definitions from various theorists, and then identified indicators for each category that characterized how students in the study would communicate. He found evidence for HOT in 70 percent of talk coded from transcripts of videotapes. (Note:

need to change summarization paragraph above to mention this is not online.) The high percentage related to other studies discussed in this review might relate to the data evaluated, transcripts of videotaped interaction with the tool versus distributed online discussions.

Research related to individual engagement in critical thinking online (Perkins 06, Jacob 08), also blended models and studied the domain for indicators. Perkins created a simple model consisting of 4 phases: clarification, assessment, inference, and strategies. While Garrison's practical inquiry model was considered it was not chosen as the sole framework because of the author's purpose of studying individual engagement in critical thinking not in the online community as a whole. Despite the research focus, however, results were presented on group and individual levels. For the group studied discussion focused mainly on clarification and least on strategies. Examples were presented showing how the model can be used to compare individual students.

In a study of critical thinking in an undergraduate online math problem discussion board (Jacob 08), the Perkins model was used as a base and then indicators were developed from the domain of math problem solving and then qualified using Paul and Elder's qualities of an accomplished critical thinker.. Like Perkins the authors found on a group level most postings were in the lower levels of HOT -- clarification and assessment -- while individual-level analysis showed varying levels of HOT engagement within categories.

Other examples of a multi-model approach were 05 Mclean and 04 Newman. As Jacob 08 did, McLean used Paul and Elder's intellectual standards for critical thinking: clarity, accuracy, precision, relevance, depth, logic, and breadth as part of a mixed-model approach to the analysis of critical thinking quality in an undergraduate correspondence psychology course. Results

showed that the quality of critical thinking was in line with other studies of critical thinking in undergraduate students, low to moderate. Critical thinking areas that were not displayed often included “supporting inferences and interpretations” and “making value judgements.”

Newman was also interested in measuring quality but also concerned with how HOTS was demonstrated in online and face-to-face discussions. He studied students examining controversial issues in IT and society. For this he combined models of Henri (is this the right author or is it multiple), who identified criteria on which to evaluate the cognitive dimensions of CMC, and Garrison because, in his opinion, Garrison’s model aligns most closely to Henri’s and relates to the social meaning construction that occurs in group critical thinking processes. However, to factor in a quality measure Henri’s model was also used to measure depth of cognition leveraging a list created by Henri of paired opposites that describe surface processing and deep processing indicators. The authors used that as a starting point to create their own paired indicators, such as “Offering judgments or solutions without explanations or justification” and “Justifying solutions or judgments.” Results from analyzing tape-recorded seminars and computer conference transcripts showed similar ratios for critical thinking in both face-to-face and computer conferencing. However, computer conferences showed more information linking behavior and less generation of new ideas. The authors point to the computer conferencing system used that made it easier to review previous messages before responding, making it easier to integrate ideas than in a face-to-face discussion.

Meyer also investigated the differences between face-to-face and online discussions with regard to demonstration of HOTS skills among graduate students in an educational leadership class. However to determine the differences a student questionnaire was given instead of conducting a content analysis of tape recordings of face-to-face seminars. This reflects the difference in what

the authors were measuring with regard to comparing the mediums, quality of critical thinking versus favorability for learning. Meyer used the PI model solely during the content analysis and found similar results to Garrison when he employed the model, a low number of resolution-type postings. In line with Newman's findings that new ideas were generated more often in brainstorming, Meyer reported that certain types of activities, such as brainstorming and topics benefitting from high energy and excitement, appeared to work better face to face. Likewise, it was reported that topics requiring reflection may work better in an online discussion, which reflects Newman's findings of more integration in the online setting.

While Schrire did not use a multi-model approach, he did examine the viability of different models for categorizing individual and socially distributed cognition. He used interaction pattern maps combined with three models of cognition—Bloom's taxonomy, SOLO taxonomy, and the Practical Inquiry model -- to investigate the relationship between individual and socially distributed cognition among doctoral students using asynchronous online computer conferencing in a technology in education class; he considered Bloom's and Solo to be individual cognition models. In his analysis he found a number of different message patterns including scattered, message chains, and instructor centered. When comparing categorizations among the three different models roughly half to two-thirds of the messages showed evidence of higher order thinking. In general, for individual cognition the models performed similarly. However, the practical inquiry model may be more apt to identify cognitive processes in group contexts as there appeared a stronger relationship between interaction patterns and cognition than with the SOLO and Bloom taxonomies. For instance ...

While individual and social cognition was a consideration by some in the choice of framework, another factor considered was applied versus abstract definition of cognition. For instance,

Jacob noted that the educational definition of problem solving did not consider ill-structured problems that occur in real-life and therefore was not suitable. Garrison when addressing possible issues with his own results from applying the Practical Inquiry model mentioned its appropriateness in that it reflected real-life situations versus other models that were "...based on abstract logical thinking processes, such as deductive thinking and analysis of arguments, with little consideration of critical discourse." And, in the process of trying to apply Bloom's taxonomy for identifying skill levels of undergraduates in software engineering and experienced software engineers, it was reported that as it stood it was not relevant given that Bloom's was focused on scientific process not operational as in SE and that it was never intended to be used outside teaching.

### **Units of Analysis**

Across the papers various units of analysis were used. For instance, Schrire in his case study approach used multiple units of analysis from examining the computer conference as a whole to the discussion thread to messages to speech segment and used interaction pattern mapping to help identify threads of interest. Herrington and Oliver examined various methods for identifying the appropriate unit of analysis considering analysis by individual words, categorization by passage, each student utterance, and unit of meaning. The individual words approach didn't apply since the task was to identify themes. Passage categorization was abandoned because of difficulty identifying when one passage ended and another began. Utterance wasn't used because at times various turns of talk included multiple category types which made simple counting of the most important category problematic because it may be misleading. Unit of meaning, or each instance of a type of talk, was the approach decided upon because it captured types of talk that may have been missed by the other methods. Perkins and

Jacob followed the same approach and in the case of possible multiple categorizations took the predominate or most important meaning. Newman also coded by unit of meaning but counted only obvious examples, avoiding grey areas to make the assessment task easier, but allowed the coding of multiple categories in a unit as did McLean.

Meyer, who used the Practical Inquiry framework, used the same unit of analysis as Garrison did when he tested his framework, coding a student's entire message in a thread. Garrison's rationale for using the message as the coding unit was based on a number of reasons. First, it was clearly identifiable by the coder, unlike other units, even sentence-level units. Second, sub-message units make the process more onerous as each coder must decide on it. Lastly, the message content and length is determined by the author not the coder. When multiple phases or contradictory phases were encountered, Garrison used a process of coding down when the phase was unclear and coding up when multiple phases were present.

Noting that Garrison had not intended for the PI model to be "immutable" and lack of further testing of the model since first introduced, Fahy looked to apply and refine the PI model. In addition, he wanted to compare results from applying the PI and Transcript Analysis Tool (TAT) for identifying among other things cognitive presence. The choice to make a comparison was based on concerns that the PI model was not suitably complex, noting that compared to transcript analysis studies in general the PI model's 4 phased approach was seen as "relatively low"; the message unit level of analysis and the coding up and coding down process used. In addition, Fahy notes that Garrison traded off accuracy for reliability when Garrison suggests, "Submessage level units [ie, sentences] may be introduced in future confirmatory studies if increased precision is warranted."

The TAT was adopted to address the concern of accuracy although while trading off simplicity and reliability. It involves coding at the sentence level into one of 8 categories, of which five are major categories: 1) questions (horizontal or vertical), 2) statements (referential or non-referential), 3) reflections, 4) scaffolding comments, or 5) paraphrases and citations. Sentences can be coded into one or more categories. In making the comparison it was expected that each tool would show facets of behavior that the other did not. In this subject population, students in a graduate credit online course, the proportions in the given phases of the PI model were similar to the original study proportion, with exploration accounting for a substantial portion of the messages. What the TAT model did show was insight into the other category of the PI model, revealing social and network functions to be a large part of this category. Also, trigger postings tended to have fewer sentences than exploration and resolution, as one might expect. Lastly, when analysis is performed at sentence level it may show communicative strategies used during critical thinking as part of the community of inquiry.

### **Exogenous factors affecting exercise and measurement**

When Garrison conducted the initial pilot of the PI model he was struck by the low proportion of responses in the integration and resolution categories. In his discussion he offered that possibly the model was not appropriate, but also that the instructional design and facilitation of the course and the computer conferencing tool used could also be factors. Meyer in noting similar results to Garrison expands on that list to include the complexity of the issue and student skill in formulating and testing resolutions.

These factors point to the instructional environment and student predispositions as factors that influence the exercise of HOT behaviors online. The measurement is further compounded by the difficulty of trying to assess individual thought processes, which is an inherently subjective

process, and the extent to which students make cognitive presence available in online discussions. As such, Garrison recommended using triangulated measures to supplement the transcript to provide more accuracy to the interpretation.

---

Chan – although not dealing with online students had some good points about using the various taxonomies together to deal with ambiguity. Has research to back up.

Herrington- I believe said to also consider your domain and tailor the frameworks.

Newman (makes nice summarization why Ennis doesn't work and why Garrison, which the author says aligns to Henri's cognitive dimension of CMC, is the better way to go.- said to use Garrison for online learning because the environment is situated learning. The other frameworks relate more to individual cognitive. This

[http://members.iinet.net.au/~aamcarthur/11\\_March\\_2008\\_files/Learning\\_Theories\\_of\\_Educational\\_Design.pdf](http://members.iinet.net.au/~aamcarthur/11_March_2008_files/Learning_Theories_of_Educational_Design.pdf) has a nice theory development continuum image that can express that thought.

Schrire in her study found relationship between PI model (Garrison) and representing relationship between cognition and interaction.

Formal versus informal models of CT – Mclean mentions Ennis might be more related to formal than Henri's and Garrison discusses the "applied" nature of his framework.

Derived indicators from domain

---

Halpern

Questions in for me in general:

1)When would I use a social interaction model such as Garrison and an individual cognition model such as Bloom? Could very well be that certain tasks in OSS are not necessarily facilitated by social cognition, but would those tasks still be social cognition if used artifacts created by the community to figure out? So, in essence, is there a move from less socially facilitated knowledge construction to more social knowledge construction as you move through the community and the tasks you undertake? What does that mean for the type of cognition employed in earlier tasks? Furthermore, Azuma 04 mentions that when he tried to tailor Bloom's to SE, found that some higher levels of Bloom's may not apply to some tasks. This also adds credence to the idea that HOT skills may not get employed in beginning tasks in OSS.

This leads to some thoughts:

2)What is the relationship to social knowledge construction, ill-structured tasks, and HOT thinking?

### **Choice of Framework (trying to get at construct validity)**

In deciding on a framework to use for organize indicators of HOT behavior there were two major considerations. First, was I interested in the extent to which the group exhibited HOT or individual member behaviors or both? Second, was what the intended purpose of the creation of the taxonomy and did it seem reasonable to modify for my purposes. Given the plethora of taxonomies, I felt it important to narrow down to a few that seemed reasonable and start from there, at least to get a feel of how well they mapped to the domain I'd be studying. If after trial the framework didn't apply I'd revisit using one I'd abandoned.

**Case study of application to experience as contributor to AnkiDroid**

**Case study of application to OpenMRS mailing list**

## References