Using the Holistic Critical Thinking Scoring Rubric to Train the Discovery of Evidence of Critical Thinking

Peter A. Facione

Background
Sometimes the best learning experiences come from efforts to teach with a mind toward evaluation. Consideration of how one might measure the achievement of learning goals forces one to develop a clear understanding of what evidence of learning should be observable in the classroom or clinic.

Noreen Facione and I developed the Holistic Critical Thinking Scoring Rubric (HCTSR) in 1994 in response to requests for a tool which (a) could be used to evaluate a variety of educational work products including essays, presentations, and demonstrations, and (b) works as both a pedagogical device to guide people to know what critical thinking is as well as an assessment tool useful for both formative and summative purposes. The HCTSR integrates both the skills and the dispositions involved in critical thinking. It has been translated into many languages and republished in numerous texts on the teaching of critical thinking. We attribute this rubric’s wide popularity with educators across many disciplines and professional fields to its ability to describe levels of quality in reasoning using common language. Another rubric we’ve developed is the Professional Judgment Rating Form (PJRF), intended for use in rating demonstrated critical thinking skills and disposition in simulation labs and clinical placement settings.

Note: Chapter 1 of Think Critically, (Facione & Gittens, Pearson Education) teaches beginning college students how to use the HCTSR developmentally. The chapter includes sets of sample essays evaluated using the HCTSR, and exercises for students to further develop their understanding and skill with the tool.

How I use the HCTSR
One of my favorite lessons with the HCTSR is taught not to students but to experienced faculty colleagues. Typically at a professional conference workshop on learning outcomes assessment I use it in an exercise intended to demonstrate how rubrics are developed and used. But more importantly, I use it as a way of helping faculty to identify evidence of a thinking process embedded in a piece of written work or in an oral presentation. If we are to teach fellow clinicians or health science students how to better approach clinical problems, it is imperative that we train ourselves to be better
at hearing the thinking of those we mentor and skilled at helping them to analyze their thinking for its quality. The HCTSR can be used by anyone to rate thinking, but the ratings will be flawed unless the rater can identify evidence of strong or weak critical thinking.

The exercise is also useful for showing the importance of establishing inter-rater reliability as a precondition for the valid and reliable use of rubrics as summative assessment tools. In this case the inter rater reliability rests on the group’s ability to identify evidence of strong or weak critical thinking skill or disposition.

The demonstration goes as follows: We read the HCTSR together talking about the construct of “critical thinking” so that we all understand at an abstract level what this tool is intended to measure. We talk about disciplining ourselves to respond to the critical thinking manifested in the student work, and not other important but separate features of that work product, such as its factual correctness, its grammar or its writing quality. Again, while these elements are essential for an overall evaluation of a student’s project, at the moment our focus is on the critical thinking involved. Typically everyone in the workshop, experienced faculty one and all, understands the distinctions that need to be made and promises to focus only on the critical thinking when they apply the HCTSR to the actual student work.

That understood, I distribute a small number of short essays written by students on some controversial topic, e.g. genetic modification of foods. Any piece of student work that requires them to discuss a decision making process can be used for this exercise as it could reasonably be expected to contain evidence of the student’s thinking process. I use essays that were written spontaneously, meaning that the students did not have the chance to do any preparation or gather any data prior to being asked, in class, to spend 20 minutes writing their opinions and explaining their reasons. The faculty in the workshop are provided with a description of the essay writing assignment and invited to rate the essays using the HCTSR (see below). After a suitable period of time, the great majority have given a score: 4, 3, 2 or 1, to each of the essays. A rating of 3 is a good score, and a 4 is a very good score. Similarly a 2 is a poor score, but 1 is far worse. We then tally the scores for each essay, and inevitably we find that on some essays there is reasonable consensus. For instance, all or most of the faculty have given an essay a positive score (4 or 3) or a negative score (2 or 1). But, and I set it up this way intentionally, there is always at least one essay where the rating is not so obvious and we discover that the group has assigned the full spectrum of scores, with the majority of scores being 3’s or 2’s.

We then debrief the situation together. Splitting the ratings more or less equally across ‘good’ and ‘poor’ scores is not a good situation, obviously. If this had been a “pass-fail” assessment, which rating would it have been? Pass or fail? If the student had been lucky (unlucky) enough to have registered in a section taught by one professor she would have been given a passing grade on that assignment, but the same assignment would have earned a failing grade by a colleague professor teaching another section of the same course. How can we, as a profession, make sense of such a demonstrably unwarranted variance in grading of the same work? And the questions go on from there. How would we adjudicate a dispute between professor and student over a grade which seems on the evidence to be capricious or arbitrary? How would we advise the student to improve his or her work, when in fact the work stands as “excellent” in the eyes of some of us and as “failing” in the eyes of others of us?

To overcome our problem with divergent ratings, we discuss the reasons why the ratings were assigned. First the faculty talk together in pairs, then in groups of four, then the fours combine to make groups of eight. At each level of greater size, the faculty teams are asked to explain what they saw in the various student essays that led them to give it a particular score on the HCTSR. This is a critical thinking exercise in itself, where raters must reanalyze the essays, identifying the evidence justifying their judgment and explaining it to the other raters. It calls on the critical thinking skills of interpretation, analysis, evaluation, and explanation. It also calls for each participant to give fair-minded attention to the reasons others have for their divergent ratings, and to be willing to reconsider in light of the analyses being offered by others about the essay. If the faculty raters have strong critical thinking dispositions – and if they have not dominance structured, that is “dug in,” around their preferred opinion – they can often hear what their colleagues are saying and at times be persuaded to change their scores. The practice of ‘reason giving’ is one that I
hope they will generalize to other professional and life situations. Outliers may emerge, but in general there is a
tendency toward consensus; this is particularly true when the faculty have generally evaluated a given essay as positive
(4 or 3) or as negative (2 or 1). In this case, they usually gravitate toward one of the scores in each pair as their groups
come to consensus regarding just how good or how bad the essay was.

The essay with ratings that range all the way from 4 to 1 at times defies ready consensus. Fundamentally the faculty
have difficulty deciding simply whether it represents good critical thinking or poor critical thinking. This is in part
because (a) the raters are still relatively inexperienced in talking with colleagues about their rating in any way
whateover, or (b) they initially have difficulty explaining which specific aspects of a given piece of student work led
them to form the judgment they have made. But the primary reason faculty disagree is because of (c) confounding
critical thinking with other factors such as rhetorical style or writing quality. Students who use a lot of irony, sarcasm,
or humor in writing their opinions seem to engender more
polar responses in the faculty. And while it may be
inappropriate for a variety of reasons to address the topic of genetic alteration of food with a Dennis Miller style rant
in an academic context, that approach in and of itself does not make the critical thinking evidenced in the writing
stronger or weaker. One must still make a rating of the rhetorical rant using the quality criteria outlined in the HCTSR.
Similarly, poor grammar or composition must be ignored if the quality of the thinking being evidenced is strong. Thus,
being able to parse out with precision those elements of the students’ work that should or should not appropriately be
considered when using the HCTSR, or in any other “grading” context, becomes a chief learning outcome of this lesson.

A second learning outcome I have for this exercise if that the faculty will come to understand that applying a rubric is
like applying other abstract principles - such as promotion and tenure guidelines – we all understand what the guidelines
mean by reference not only to the guidelines themselves but also by reference to the accumulated prior precedents.
That is why sharing with students one’s grading standards, as well as examples of the graded work of prior students,
is an effective way to help the students understand where you “set the bar” for successful vs. unsuccessful work. With
practiced use rubrics are internalized and, in the case of the HCTSR, we are soon able to identify a piece of work as
characteristic of a ‘4,’ quite distinct from another piece of work that would be rated a ‘2.’ And more importantly, we
become practiced at identifying significant evidence of the thinking process found in the essay and of providing our
reasons for assigning the specific rating.

The third learning outcome intended by this workshop exercise with experienced colleagues is that they should come
to appreciate the importance and the value of continuing collegial conversations regarding how the assignment of
ratings and the evaluation of practice behavior requires the responsible explanation of how standards and guidelines
are being applied in the classroom and the workplace. This becomes not only a matter of excellence in professional
practice, but also a matter of justice and departmental responsibility and liability. In my judgment, to know that we
disagree as faculty teaching alternative sections of the same course on whether a given piece of student work is a “pass”
or a “fail” and not to address that disagreement in an objective, professional, informed, and fair-minded way is nothing
short of irresponsible.

Feedback from Participants
I used this exercise with student work and the HCTST in workshops with faculty for over fifteen years and the reaction
is always positive. Done with humor and respect, and with a willingness on all of our parts to “let the chips fall where
they may,” this exercise turns out to be both educational and enjoyable. I think it would be a more difficult exercise to
use if the faculty were all from the same department, rather than the typical mix of people we find at national meetings.
For one, they would have histories with one another, complex interpersonal and organizational relationships which
began before the workshop and which must endure after the workshop. The lesson would have to be adapted, more
precautions taken to assure that people can participate with minimal perceived risk to their long term collegial
relationships, and more awareness that any given observation or comment comes with a potentially substantial amount
of unspoken baggage. But, forty years in higher education, and much of that as chair, dean and provost, helps sensitize one to such things. And good colleagues realize that we should be able to talk in a fair-minded and collegially respectful way about our grading and performance ratings, not only for the sake of the health of our departments and programs, but also for the sake of the better educational development of our students and colleagues.

**Using the Holistic Critical Thinking Scoring Rubric**

1. **Understand the construct:** This four level rubric treats critical thinking as a set of cognitive skills supported by certain personal dispositions. To reach a judicious, purposive judgment a good critical thinker engages in analysis, interpretation, evaluation, inference, explanation, and meta-cognitive self-regulation. The disposition to pursue fair-mindedly and open-mindedly the reasons and evidence wherever they lead is crucial to reaching sound, objective decisions and resolutions to complex, ill-structured problems. So are the other critical thinking dispositions, such as systematicity, reasoning self-confidence, cognitive maturity, analyticity, and inquisitiveness. [For details on the articulation of this concept refer to Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. ERIC Document Number: ED 315 423.]

2. **Differentiate and Focus:** Holistic scoring requires focus. In any essay, presentation, or clinical practice setting many elements must come together for overall success: critical thinking, content knowledge, and technical skill (craftsmanship). Deficits or strengths in any of these can draw the attention of the rater. However, in scoring for any one of the three, one must attempt to focus the evaluation on that element to the exclusion of the other two.

3. **Practice, Coordinate and Reconcile:** Ideally, in a training session with other raters one will examine sample essays (videotaped presentations, whatever is being evaluated) which are paradigmatic of each of the four levels. Without prior knowledge of their level, raters will be asked to evaluate and assign ratings to these samples. After comparing these preliminary ratings, collaborative analysis with the other raters and the trainer is used to achieve consistency of expectations among those who will be involved in rating the actual cases. Training, practice, and inter-rater reliability are the keys to a high quality assessment.

Usually, two raters will evaluate each essay, or assignment, or project, or performance. If they disagree there are three possible ways that resolution can be achieved: (a) by conversation between the two raters regarding their evaluations, (b) by using an independent third rater, or (c) by taking the average of the two initial ratings. The averaging strategy is strongly discouraged. Discrepancies between raters of more than one level suggest that detailed conversations about the CT construct and about project expectations are in order. This rubric is a four level scale, half point scoring is inconsistent with its intent and conceptual structure. Further, at this point in its history, the art and science of holistic critical thinking evaluation cannot justify asserting half-level differentiations.

When working alone, or without paradigm samples, one can achieve a greater level of internal consistency by not assigning final ratings until a number of essays, projects, assignments, performances have been viewed and given preliminary ratings. Frequently natural clusters or groupings of similar quality soon come to be discernible. At that point one can be more confident in assigning a firmer critical thinking score using this four level rubric. After assigning preliminary ratings, a review of the entire set assures greater internal consistency and fairness in the final ratings.

```
*   *   *   *   *
```

Although this series focuses on health science content, the techniques are transferrable to all types of training programs and educational projects.

Download other essays in this series for valuable training techniques that focus student learning of reasoning skills and thinking mindset. See Resources on our website.
The Holistic Critical Thinking Scoring Rubric - HCTSR
A Tool for Developing and Evaluating Critical Thinking

Peter A. Facione, Ph.D. and Noreen C. Facione, Ph.D.

Strong 4. Consistently does all or almost all of the following:

- Accurately interprets evidence, statements, graphics, questions, etc.
- Identifies the most important arguments (reasons and claims) pro and con.
- Thoughtfully analyzes and evaluates major alternative points of view.
- Draws warranted, judicious, non-fallacious conclusions.
- Justifies key results and procedures, explains assumptions and reasons.
- Fair-mindedly follows where evidence and reasons lead.

Acceptable 3. Does most or many of the following:

- Accurately interprets evidence, statements, graphics, questions, etc.
- Identifies relevant arguments (reasons and claims) pro and con.
- Offers analyses and evaluations of obvious alternative points of view.
- Draws warranted, non-fallacious conclusions.
- Justifies some results or procedures, explains reasons.
- Fair-mindedly follows where evidence and reasons lead.

Unacceptable 2. Does most or many of the following:

- Misinterprets evidence, statements, graphics, questions, etc.
- Fails to identify strong, relevant counter-arguments.
- Ignores or superficially evaluates obvious alternative points of view.
- Draws unwarranted or fallacious conclusions.
- Justifies few results or procedures, seldom explains reasons.
- Regardless of the evidence or reasons, maintains or defends views based on self-interest or preconceptions.

Weak 1. Consistently does all or almost all of the following:

- Offers biased interpretations of evidence, statements, graphics, questions, information, or the points of view of others.
- Fails to identify or hastily dismisses strong, relevant counter-arguments.
- Ignores or superficially evaluates obvious alternative points of view.
- Argues using fallacious or irrelevant reasons, and unwarranted claims.
- Does not justify results or procedures, nor explain reasons.
- Regardless of the evidence or reasons, maintains or defends views based on self-interest or preconceptions.
- Exhibits closed-mindedness or hostility to reason.

(c) 1994, 2009, Peter A. Facione, Noreen C. Facione, and Measured Responses LLC, Hermosa Beach, CA USA
Published by The California Academic Press / Insight Assessment, Milpitas, CA 95035

To download free copies of this rubric visit http://www.insightassessment.com/HCTSR.html

Permission is granted to students, faculty, staff, or administrators at public or private educational institutions and organizations for unlimited duplication and free distribution of the Holistic Critical Thinking Scoring Rubric for teaching, assessment, research, or other non-commercial uses, provided that no part of the scoring rubric is altered, that the copies are distributed free of charge, and that “Facione and Facione” are cited as authors.

www.insightassessment.com USA Phone: (650) 697-5628
03/4/09 (Ver 2.50.00)
For permission to include this rubric in materials that will be sold, contact Insight Assessment.