

FINAL REPORT

Researching Assessment Methods in Tutorial Education

Robert J. Beck, William F. Skinner, & Lynsey A. Schwabrow

Lawrence University

Executive Summary

The purpose of this study was to research and develop alternative assessment methods for tutorial courses in liberal arts education. Nine faculty members from Lawrence University, Macalester College, The College of Wooster and Williams College developed a Shared Assessment Method (SAM) for assessing three groups of traits: Independent Thinker (*takes teacher roles; inquiring mind; self-assessment; learning to argue*); Intellectual Maturity (*ability to work with complexity/uncertainty; takes intellectual chances*); and Creativity (*idea generation; curiosity; takes multiple perspectives; connects disparate information and ideas*). In 21 tutorial courses involving 40 students the faculty tested rubrics assessing the traits through observations, interviews and rating scales. The results of the study as analyzed from faculty case studies and rubrics suggest that the selected traits were valid and important. The conclusion that the traits are valid for purposes of assessment, and enhanced the pedagogy of tutorials is based on the following findings:

- Faculty and students were able to observe and measure the traits during two cycles of tutorial courses;
- Faculty found methods for encouraging student development of the traits through course plans, assignments and teaching strategies;
- The rubrics discriminated different levels of traits in stronger and weaker students;
- Mean levels of student self-estimated and faculty-estimated levels of all traits rose during the tutorials. Statistically significant improvement in the traits from baseline to final assessment periods were recorded for all the traits of Independent Thinker and Intellectual Maturity and the Creativity traits of *idea generation* and *connectivity*;
- Student evaluations indicated that they integrated the traits into their learning goals;
- Several faculty proposed using the rubrics to assess college-wide tutorial programs;
- In some cases, faculty noted that students of equal grade point averages performed unequally on the trait assessments, suggesting that the rubrics captured qualities above and beyond content mastery;
- Trait validity findings stand up across tutorials offered by a wide range of faculty representing disciplines in the Arts, Humanities, and Sciences.

Nonetheless, several faculty concluded that student *curiosity* (Creativity) was already at high levels at the start of courses and no important strategies emerged for supporting this trait. Other tutors found too much overlap between *connectivity* and *multiple perspectives* (Creativity). We

propose reducing the number of traits to eight: Independent Thinker (4); Creativity (2- *idea generation, takes multiple perspectives and connects disparate information and ideas*); Intellectual Maturity (2). There were no significant proposals to add new traits.

Going forward, we recommend applying the trait rubrics to assessments of tutorials (1-3 students), hybrid courses where tutorials are additions to conventional small courses like seminars (5-10 students), and, perhaps, liberal arts education as a whole. Because The College of Wooster already uses their independent study courses as culminating projects, they essentially serve as assessments for the full undergraduate education. Lawrence University is now introducing the senior experience as a culminating project for students. Both of these examples refer to tutorial-like programs. The rubrics, therefore, would be useful for assessing a student's liberal arts education. To reduce faculty workload the traits would be used in syllabi to define goals, but only students would fill out rating scales, four or more times per term or semester, and orientation and debriefing class sessions would be retained to teach the criteria and assess effectiveness. The faculty in this study are important resources to train and supervise other faculty on each campus in the extension of use of these assessment/pedagogic rubrics.

Introduction

The purpose of this study was to research and develop alternative assessment methods for tutorial courses in liberal arts education. The study responded to recent calls by several prominent educators (Conner, 2007; Bok, 2006; Katz, 2008) for new methods of assessment in liberal arts colleges. In 2008, the Association of American Colleges and Universities urged that colleges develop quality standards themselves and not rely on external agencies. Many colleges have answered this criticism by undertaking formal and institutional research projects to determine whether and how undergraduates developed socially/intellectually from their entry to graduation. A total of 9 faculty and 40 tutorial and 8 non-tutorial students constituted the study population. The faculty were drawn from four liberal arts colleges representing the natural sciences, social sciences, humanities, and fine arts: Lawrence University, Macalester College, The College of Wooster, and Williams College. The faculty participated in the study as co-investigators using tutorial-type courses as research settings.

The group developed a set of student learning outcomes and rubrics for rating traits and observational methods during two cycles of tutorial courses. A description of the process used to develop the first set of learning outcomes and associated traits (SAM-1) appears in Appendix 2 along with the results of the rubric analysis (Appendices 3, 4, and 5). What follows is the methodology for the second iteration of the Shared Assessment Method (SAM-2).

Method

Development of Learning Outcome Rubrics: SAM-2

Following the first year of the study, seven of the eight original faculty and one new faculty participant came together for a Conference on Tutorial Education and Assessment in fall 2009 featuring keynote speaker Charles Blaich, Director of Inquiries from the Center for Inquiry in the Liberal Arts at Wabash College. Faculty actively participated by summarizing their experience with implementing the rubrics during SAM-1 and provided their recommendations going forward.

Based on SAM-1 rubric analyses and case studies completed by faculty, all participants discussed desired changes for the learning outcome rubrics that would be used for SAM-2. As each trait was described, the faculty assessed their viability and offered evidence that they had previously observed such traits in their tutorials. To further validate the learning outcomes, the

rubrics were assessed by faculty for importance and scalability. Were the outcomes capable of being rated and observed by faculty and students alike? Faculty should be able to assess that a given student has exhibited or has not exhibited the trait in question and that he or she has an estimable level of the trait at certain points during the course of a tutorial; a student should be able to self-assess that (s)he has exhibited or has not exhibited the trait and has an estimable level of the trait at certain points during the course of a tutorial. As an outcome of this discussion, faculty accepted certain learning outcome traits and rejected others. Changes included the elimination of traits to reduce the overall number, and the descriptions of the chosen traits were revised. SAM-2 rubrics ultimately included 10 traits within the three broad groups of learning outcomes of Independent Thinker, Intellectual Maturity, and Creativity:

Learning Outcome	Traits	Description
INDEPENDENT THINKER	Independence	the ability to take teacher roles by setting topics, asking questions to originate new topics, summarizing discussion and assessing ideas
	Developing an inquiring mind	asks high-level questions, such as why-questions
	Acquiring self-assessment skills	self-assesses one's claims and arguments; the ability to recognize biases in thinking and predispositions to make certain judgments, shares or communicates critique of one's biases and prejudgments
	Learning to argue	argues effectively by making conceptual claims backed with supporting theory, reasoning and evidence; does not imply emotional and contentious argument
INTELLECTUAL MATURITY	Complexity/Uncertainty	the ability to work with complex problems, issues, and information
	Takes intellectual chances	willingness to state positions and arguments without worry of saying something wrong, making mistakes, or risking failure
CREATIVITY	Idea generation	generates new ideas, variations of or alternatives to solving problems, a novel way of analyzing or re-conceptualizing a topic or idea in the context of what the student knows and understands, interesting and creative restatements of others' ideas, unusual ideas, interesting theories
	Curiosity	the desire to learn or know more, ability to become absorbed in the topic, discovers a new line of inquiry or question of a topic and wishes to persist and sustain in exploring the topic
	Multiple perspectives in problem solving	sees a problem from multiple perspectives, compares and contrasts approaches, uses multiple disciplines
	Connectivity	ability to bring together or synthesize disparate bits of information, makes connections between already established ideas or theories, connects disciplines

Tutorial and Non-Tutorial Course Organization and Evaluation: SAM-2

Tutorial courses enrolled 1-3 students to write papers, engage in discussions, and, if required, present their work. A total of 20 tutorial students and 8 non-tutorial students participated in SAM-2. Learning outcome rubrics were used for the assessment of student

performance across discussions, papers, and/or presentations at three times during the tutorial. These three assessment periods were referred to as the baseline, midpoint, and final assessments. Students were also instructed to complete the rubrics to self-assess themselves during assessment periods. Each trait was evaluated on a 5-point scale (1=Never to 5=Very Frequently) concerning how often the student exhibited that trait. Faculty were also given the opportunity to record detailed qualitative observations they made of the traits exhibited in their tutorial student(s). During SAM-2 faculty were encouraged to record additional traits on the rubrics that they observed or fostered during their tutorial or non-tutorial courses. Each assessment period was audio recorded and transcripts were created.

In addition to tutorial course assessment, SAM-2 included assessment of one or more students from a non-tutorial course who were chosen at random by faculty. The assessment periods of the students' work were identical to the tutorial course evaluations. Only instructors completed learning outcome rubrics and no recordings were made with the students in non-tutorial courses.

The following procedural timeline indicates the evaluation points during each tutorial and non-tutorial course:

- Orientation interview – to describe the goals and methods of the course, review the informed consent form including compensation for the student. Discussion and use of the learning outcome rubrics to be used in assessing student performance. This interview was not recorded.
- Baseline assessment – occurs during the second or third week of class and serves as the baseline measure of the learning outcomes. The rubrics are used in assessing the discussion that occurs in one class and any written work that occurs during this window of time. This discussion was recorded and transcribed.
- Midpoint assessment – occurs during the fifth (for terms) to eighth (for semesters) week of class and serves as the mid-point measure of outcomes. The same procedures are followed as for the baseline assessment.
- Final assessment – occurs during the last couple weeks of the tutorial and serves as the end-point measure of the learning outcomes. The same procedures are followed as for the baseline assessment.

- Exit interview – to discuss progress and development during the tutorial. This interview was recorded and transcribed.

During SAM-1, we sought the additional validation of the traits through the correlation and triangulation of both self-report rating scales and more objective evidence, such as textual passages found in student essays and transcripts of tutorial discussions. During SAM-2 the instructors more specifically made the achievement of the traits a goal of the tutorials as well as to assess student performance. The faculty more deliberately used the traits, such as *inquiring mind*, through modeling (e.g., demonstrating their own inquiring mind), questioning (e.g., prompting students to inquire) and used *inquiring mind* criteria in annotating/correcting student essays.

The traits comprising each learning outcome were validated by a final case study written by each faculty member at the conclusion of each year of the study, drawing on faculty and student self-report ratings, course observations and analyses of student performance and work, and transcripts from assessment and interview periods.

Participants

Rob Neilson, Associate Professor of Art, Lawrence University

Ron Peck, Assistant Professor of Biology, Lawrence University

Jerald Podair, Professor of History, Lawrence University

Claudena Skran, Associate Professor of Government, Lawrence University

Nancy Grace, Professor of English, The College of Wooster

Pamela Pierce, Associate Professor of Mathematical Sciences, The College of Wooster

Patrick Schmidt, Assistant Professor of Political Science, Macalester College

Christopher Nugent, Assistant Professor of Chinese, Williams College

Stefanie Solum, Associate Professor of Art, Williams College

Research Questions

Q1. Were the selected traits validated during the two tutorials conducted by faculty?

Q2. Were the selected traits of equal importance in the tutorial context?

Q3. Did the use of selected traits as outcomes and rubrics enhance the pedagogy in tutorial courses from the perspective of faculty and students?

Q4. Did the use of selected assessment traits of independence inform the design of assessments of other non-tutorial courses?

SAM-2 Case Studies Analysis

The results of the study suggest that the selected learning outcomes of Independent Thinker, Intellectual Maturity and Creativity are valid and should be applied to assessments of tutorials (1-3 students), hybrid courses where tutorials are additions to conventional small courses like seminars (5-10 students), and, perhaps, liberal arts education as a whole. The conclusion that the traits are valid is based on several findings: a) faculty and students were able to observe and measure the traits during two iterations of tutorial courses; b) faculty found methods for encouraging student development of the traits through course plans, assignments and teaching strategies; c) the rubrics discriminated different levels of traits in stronger and weaker students; d) in the aggregate students levels of all traits rose; e) student evaluations indicated that they integrated the traits into their learning goals; and f) several faculty proposed using the rubrics to assess college-wide tutorial programs. In the results following we quote from the faculty case studies. A larger selection of excerpts appears in Appendix 1.

The 10 traits were generally described as “crucial”, “extremely important” and “foundational” to outcomes in liberal arts education. There were a few recommendations to consolidate some Intellectual Maturity and Creativity traits but no one suggested eliminating any of the three groups. Some individual faculty members judged that a few of the traits were less relevant to their particular courses.

Some faculty used traits to envision the goals of the course at several levels: how courses were conceived; course design; and teaching strategies. They offered that the rubrics expanded their ideas concerning the potential for student development. Most faculty found that peer interactions (2-3 students), a characteristic of all the tutorials, brought the traits into greater visibility and enabled professors to work with the traits. Other faculty found that use of the traits “freed them up” from constantly evaluating through conventional letter grading.

During SAM-2, in contrast to SAM-1, the faculty more deliberately used the constructs pedagogically as well as for assessment purposes. For example, *inquiring mind* was integrated pedagogically through modeling (e.g., demonstrating their own inquiring mind), questioning (e.g., prompting students to inquire) and in annotating/correcting student essays. As in the

learning outcome validation phase students and faculty periodically filled out rubrics to assess levels of each trait. At a *tutorial conception level* the following quotes from case studies reveal how the traits and rubrics changed faculty thinking:

Working with SAM has absolutely realigned this perspective of mine. I am able to envision the goals of this tutorial (and others) both more broadly and more specifically. In other words, I understand the potential for student development to be well beyond the parameters I first envisioned, yet I have a much more sophisticated understanding of the varied levels of experience that make for that development (Stefanie Solum, Williams)

Perhaps the most beneficial aspect of this study was having the tutorial students go through the rubric at the beginning of the year and discuss it with the instructors. The rubric led to wonderful discussions between instructors and students, where all parties were able to articulate some clear goals for the upcoming senior independent study project...I believe that this is an educationally sound practice—to give a student the rubric by which he/she will be assessed at the onset of the project, and to re-evaluate at mid-tutorial and at the end of the tutorial. If this process would take place with all tutorials, and honest feedback were to be given, then the grading process at the end of the year would not seem so mysterious (Pamela Pierce, Wooster)

At a *course design level*, the faculty was able to more or less seamlessly integrate the traits into their course plans. Faculty embedded specific traits into their assignments such as requiring students to assume teacher roles by making presentations (Independent Thinker) or ensuring that students had required readings that had a high level of *complexity*, a trait of Intellectual Maturity.

I found SAM-2 to be useful in assessing how the students progressed through the course of the tutorial and how well the tutorial succeeded as a whole. Carefully articulated learning outcomes can be a useful tool, but there needs to be a more regular way of keeping students focused on them. Filling out the rubrics after each tutorial session would be one way to do this. I feel confident that, at the very least, I will consult them (rubrics) about how to assess students in future courses, both tutorial and non-tutorial (Christopher Nugent, Williams)

In this tutorial students reported benefitting from the rubrics, so Nugent's recommendation to increase the number of ratings makes sense. For some faculty, however, filling out the rubrics was a burden. We think faculty have generally fully digested the various traits and don't need to keep filling them out. However, we suggest retaining the interviews in which tutors and students discuss how the latter are using the traits.

From the perspective of *teaching*, the faculty found that they were able to employ particular strategies to encourage the development of traits in students. These strategies were

implemented in every aspect of the instruction including the selection of works to be read and analyzed, conversations, written comments on papers, private meetings, and peer-to-peer advice. Questioning was the most ubiquitous strategy, but also teaching interventions occurred in student evaluations and in class feedback to encourage student development of the learning outcomes. Here are a few examples of tutor teaching strategies.

Independence (taking teacher role) was among the most influential traits, if not the most important. I often invoke the aphorism, “to teach is to be twice taught”, with my students, and so they have been primed to think how leadership in the classroom both evidence and solidifies one’s learning. Part of the influence of this trait is in setting the role of the tutor. Though I remain poised to enter a discussion, and do enter as needed, the trait reminds me to let there be silence. Sometimes it requires a little theatrical touch, such as looking at my notes or something as if to emphasize that I am not going to jump in on the conversation any time soon. In the vacuum, I am more likely to get students to initiate new directions. (Patrick Schmidt, Macalester)

I encouraged this trait (inquiring mind) by constantly pushing students to dig deeper into the motivations behind the arguments we read. I emphasized that understanding the arguments was only a first step. They also need to think about why particular thinkers made particular arguments....One way I encouraged the development of this trait (self-assessment skills) in the tutorial was to have students identify at least two problematic aspects of every paper they wrote...I do not feel that either students improved substantially in their ability to address their biases and predispositions on their own. At the same time, when either I, or their partner alerted them to possible biases in their arguments, they were very receptive to the idea and seemed willing to rethink things.” My primary means of encouraging another trait (generating ideas) in the tutorial is through continuously pushing students to move beyond the secondary readings to come up with their own interpretations of primary materials. (Christopher Nugent, Williams)

While strategies for both the discussion and writing were offered, most of the teaching innovations occurred during the discussion phase. Nugent’s innovation concerning writing, however, strikes us as valuable. When students review their own papers and come up with two problematic aspects, they are taking the teacher role, just as they do in presenting their papers to the group and in leading the discussion.

As reported in the case studies, the validity of the relevance of the traits in tutorials was born out in students’ responses to trait-based strategies from faculty and peers. The faculty reported that most students improved their levels of traits. These data were corroborated by positive changes in the levels of traits as analyzed in the rubric scores. Students offered additional evidence for trait validity by integrating the rubrics into their learning goals. Some students checked with tutors that their behavior indicated one trait or another. Other students

reflected on their progress on certain traits. Others specifically praised rubrics for helping them with their papers or increased abilities to self-assess and make complex arguments. Students contributed to peers' growth in selected traits by bringing in information from their own background as double majors. Peer evaluations of traits provided additional feedback.

Filling out the rubrics, especially the baseline one, seemed to make both C. and J. more aware of what they should be trying to learn and accomplish in the tutorial C. was perhaps the most self-reflective, concluding in his second assessment that "doing this" [taking chances] helps me better understand everything that we discuss. C. included the following in an observational note: 'the more we go in depth with the topic, the more I'm able to absorb it all and raise new ideas' (Claudena Skran, Lawrence)

RE: ability to take teacher roles. MA. "I like inviting audience participation through questions, but you should have a plan to lead your audience to the desired answer if nobody gets it. You can do this by restating the question, providing hints, or asking progressively simpler questions to get your audience started on the right path." PK. "You also have a natural technique for physically pointing out the things on the slide as you are describing it and drawing on the board when you need to make the point. I think that shows you are acting as a "teacher" to explain something to the audience." LY. "I could tell you made the effort to slow down and clearly describe complex concepts...you also oriented the audience to your slides and almost always provided a clear 'take-home' message for each slide" (Ron Peck, Lawrence)

From these comments it is apparent that students are well aware of what effective faculty presentations are like and they are comfortable with holding their peers to these standards, another example of practicing teacher-like behavior.

In faculty judgments, all tutorials contained stronger and weaker students. The rubrics discriminated between students who varied in performance. In many student assessments, faculty found clear patterns of high levels of traits across all Independent Thinker, Intellectual Maturity and Creativity groupings in superior students and inversely so. In some cases, faculty noted that students of equal grade point averages performed unequally on the trait assessments, suggesting that the rubrics captured qualities above and beyond content mastery.

At the onset of the project, it appeared that the rubric was motivating the weaker of my two students...it would seem natural to ask whether the rubric provides additional motivation to a more 'average' student. After a few initial meetings, R. was able to lead the discussion and initiate topics she wanted to discuss. K. on the other hand, needs a lot of guidance, and we had to prod her to keep moving (Pamela Pierce, Wooster)

RE: taking teacher role. All students showed improvement. MA. Immediately took my feedback and then refined this aspect over the course of the term. PK and LY improved, but to smaller degrees and more sporadically during the term. RE: Self-assessment. Best

students seemed to already have the ability and awareness to properly evaluate evidence, but I struggled to see or coax development of this trait in the weakest student (PK) (Ron Peck, Lawrence)

Self-assessment may be a trait that is deeply engrained in students prior to their entry in the tutorial and therefore might have had little potential for improvement. Unlike another trait, *curiosity* (Creativity), however, the tutorial offers continuous opportunities to engage in self-assessment, e.g., in writing and revising papers and in filling out rubrics. Thus, the teacher and student ratings for this trait rose over the course of the tutorial in any case.

Unresolved issues

There are a few unresolved issues about the traits. There was unanimous agreement that the traits were not useful in assessing students in most non-tutorial courses. Several reasons for this were apparent. Most non-tutorial courses do not have enough individual discussion or papers to observe behaviors that were needed to make assessments. Another perspective was that it was impossible to assess individual performance of these kinds in classes with 20 or more students. It was not apparent whether the behaviors associated with the traits are expressed but not visible to instructors in dense environments or whether the traits are simply not encouraged in courses that do not have coordinated papers, presentations and discussions. Two Williams College faculty suggested that small seminars, however, might be a good setting to employ trait rubrics.

The rubrics will have less-utility in non-tutorial courses, except as ideas to keep in mind as basic pedagogical goals. I can see the rubrics helping change, however, perhaps to develop alternatives to the graduate-school seminar model many of us continue to use for undergraduates (Stefanie Solum, Williams)

I feel extensive use of the rubrics might work best in something like a senior or junior seminar in which students all come in with some familiarity with the subject matter and have a base from which to work (Christopher Nugent, Williams)

In some cases it was hard to determine whether it was course characteristics and faculty strategies that provoked student trait development or whether students' prior knowledge and background such as, double-majors and students' entering trait levels, contributed more to trait development or failure to change.

During the course of the tutorial, some trait assessment levels for individual students failed to change. In some cases, such as for *curiosity*, particularly, those students who entered the course with high levels of the trait had little room for improvement.

SAM-2 Rubric Analysis

Tables 1 and 2 contain the results of the rubric quantitative analysis for the traits of the three groups of tutorial learning outcomes for SAM-2. The graphs show the mean rating for each trait across three time points – baseline, midpoint, and final. All of the traits associated with Independent Thinker (Table 1) showed statistically significant improvement from baseline to final as indicated by the underlined means. It is interesting to note that all the traits for Independent Thinker start at baseline between 3.0 and 3.3 and increase at the final measurement between 4.0 and 4.2. The mean rating for each trait of Intellectual Maturity also significantly improved from baseline to final. Table 2 shows that the only trait associated with Creativity that did not significantly improve between baseline and final was *curiosity*, a finding that also occurred in the first shared assessment method (SAM-1). And as was found in SAM-1, the baseline mean of 3.8 for *curiosity* is the highest starting point of all other traits. The replication of this finding confirms that students enter the tutorial with a relatively high desire to learn.

Table 3 compares the faculty and student means for the learning outcome traits in SAM-2. This analysis is based on the total number of students who completed the baseline, midpoint, and final rubrics and the corresponding faculty ratings of those students. Based on the students' self-ratings, all the traits showed improvement from baseline to final. Statistically significant improvement (indicated by superscript "a") in the traits from baseline to final occur for all the traits of Independent Thinker and Intellectual Maturity and the Creativity traits of *idea generation* and *connectivity*. As in SAM-1, there is remarkable similarity between faculty and students on the mean ratings for most of the traits. Statistically significant differences (indicated by superscript "b") in the mean ratings between faculty and student only occur at baseline for *self-assessment skills* and *learn to argue*. And it is interesting to note that as with the faculty assessment, the highest baseline assessment coming from students occurs for *curiosity* and the rating does not significantly increase by the end of the tutorial.

Table 1. Faculty Rubrics for SAM-2: Independent Thinker and Intellectual Maturity

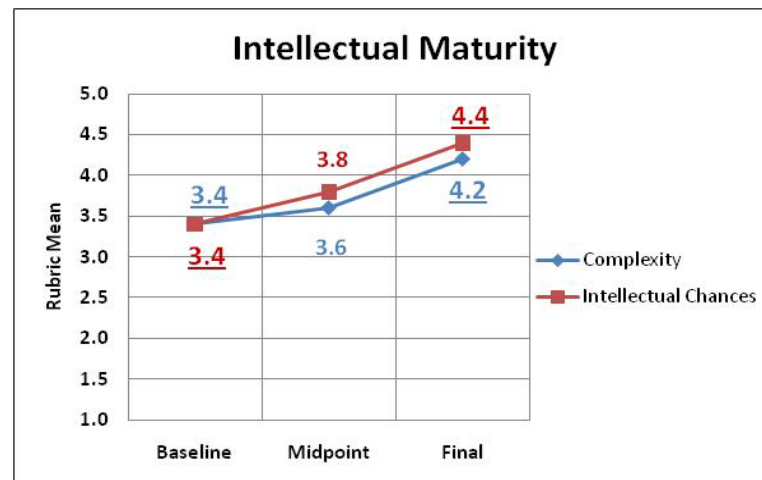
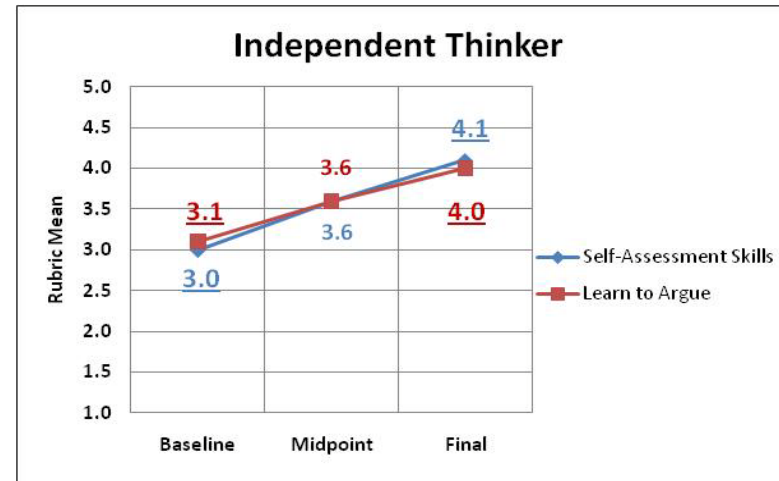
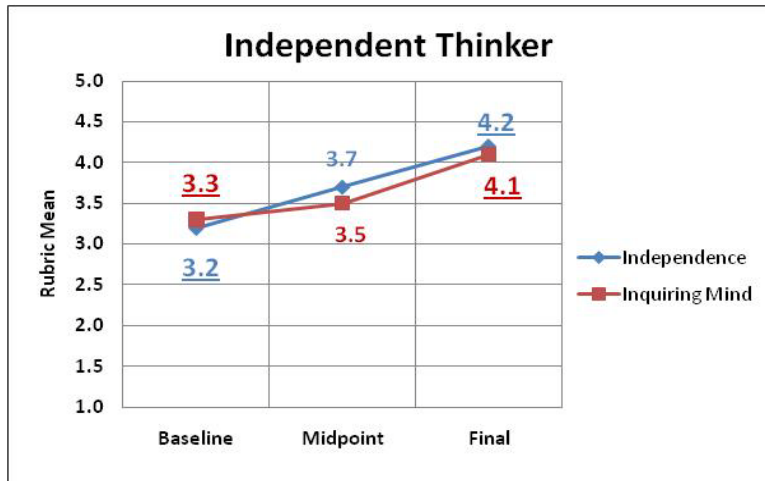


Table 2. Faculty Rubrics for SAM-2: Creativity

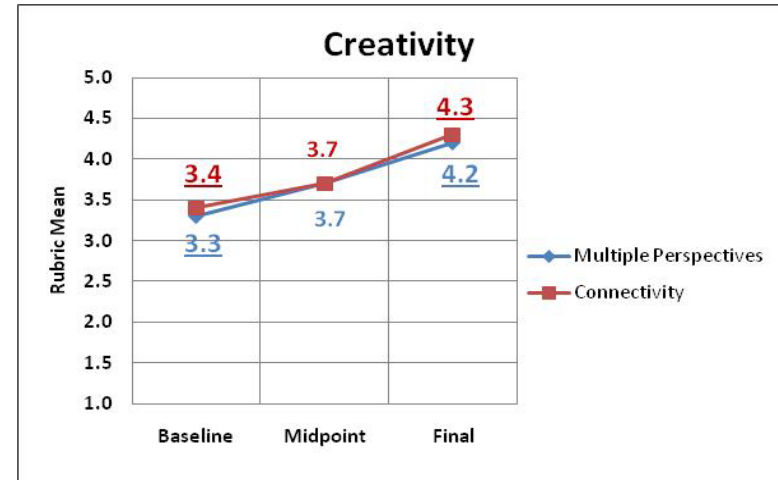
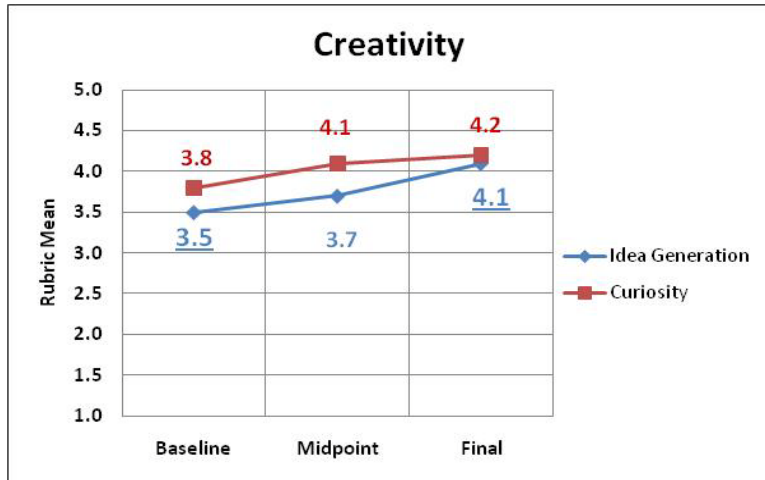


Table 3. Faculty and Student Baseline, Midpoint, and Final Rubric Means for SAM-2

Trait	Faculty Means (N=20)			Student Means (N=20)		
	Baseline	Midpoint	Final	Baseline	Midpoint	Final
INDEPENDENT THINKER						
Independence	3.2	3.7	4.2	3.3 ^a	3.9	3.9 ^a
Inquiring Mind	3.3	3.5	4.1	3.5 ^a	4.0	4.3 ^a
Self-Assessment Skills	3.0 ^b	3.6	4.1	3.8 ^{ab}	4.2	4.3 ^a
Learn to Argue	3.1 ^b	3.6	4.0	3.6 ^{ab}	3.8	4.1 ^a
INTELLECTUAL MATURITY						
Complexity	3.4	3.6	4.2	3.7 ^a	4.1	4.3 ^a
Intellectual Chances	3.4	3.8	4.4	3.7 ^a	4.0	4.3 ^a
CREATIVITY						
Idea Generation	3.5	3.7	4.1	3.3 ^a	3.9	4.1 ^a
Curiosity	3.8	4.1	4.2	4.3	4.6	4.7
Multiple Perspectives	3.3	3.7	4.2	3.7	3.9	4.1
Connectivity	3.4	3.7	4.3	3.6 ^a	4.2	4.4 ^a

^aBaseline and final student means significantly different at $p < .05$

^bBaseline faculty and student means significantly different at $p < .05$

Conclusion

The results of the study as analyzed from faculty case studies and rubrics suggest that the selected traits were valid and important.

The conclusion that the traits are valid for purposes of assessment, and enhanced the pedagogy of tutorials as based on the following findings:

- Faculty and students were able to observe and measure the traits during two cycles of tutorial courses;
- Faculty found methods for encouraging student development of the traits through course plans, assignments and teaching strategies;
- The rubrics discriminated different levels of traits in stronger and weaker students;
- Mean levels of students self-estimated and tutor estimated levels of all traits rose during the tutorials. Statistically significant improvement in the traits from baseline to final assessment periods occur for all the traits of Independent Thinker and Intellectual Maturity and the Creativity traits of *idea generation* and *connectivity*;
- Student evaluations indicated that they integrated the traits into their learning goals;
- Several faculty proposed using the rubrics to assess college-wide tutorial programs;
- In some cases, faculty noted that students of equal grade point averages performed unequally on the trait assessments, suggesting that the rubrics captured qualities above and beyond content mastery;
- Trait validity findings stand up across tutorials offered by a wide range of faculty representing disciplines of the Arts, Humanities, and Sciences.

Nonetheless, several faculty concluded that student *curiosity* (Creativity) was already at high levels at the start of courses and no important strategies emerged for supporting this trait. Other faculty found too much overlap between *connectivity* and *multiple perspectives* (Creativity) and we decided to combine these traits. We propose, therefore, to reduce the final number of traits to eight: Independent Thinker (4); Creativity (2); Intellectual Maturity (2).

Going forward, we recommend applying the trait rubrics to assessments of tutorials (1-3 students), hybrid courses where tutorials are additions to conventional small courses like seminars (5-10 students), and, perhaps, liberal arts education as a whole. Because The College of Wooster already uses their independent study courses as culminating projects, they essentially serve as assessments for the full undergraduate education. Lawrence University is now introducing the senior experience as a culminating project for students. Both of these examples refer to tutorial-like programs. The rubrics, therefore, would be useful for assessing a student's liberal arts education. To reduce faculty workload the traits would be used in syllabi to define goals, but only students would fill them out. Students should fill out rating scales at least four times per course. Orientation and debriefing class sessions would be retained to teach the criteria

and assess effectiveness. The faculty in this study are important resources to train and supervise other faculty on their respective campuses in the extension of use of these assessment/pedagogic rubrics.

References

Beck, R. J. (2007). *The Oxford Tutorial: A Learning Theory*. *Oxford Magazine*. Oxford: UK.

Beck, R. (2010). The independent thinker: Assessing student outcomes in tutorial education. Retrieved November 1, 2010 from http://www.lawrence.edu/conference/tutorials/2008/20090206_beck.pdf

Bok, D. (2006). *Our underachieving colleges*. Princeton, NJ: Princeton University Press.

Connor, W. R. (2006). "From Foxes to Hedgehogs". Center of Inquiry in the Liberal Arts at Wabash College. Podcast address given in Washington, D. C., January 27, 2006.

Generic Critical and Creative Thinking Rubric. Retrieved November 1, 2010 from http://www3.wooster.edu/teagle/docs/general_rubric.pdf

Katz, S. N. (2008). Taking the true measure of liberal education. *Chronicle of Higher Education*, p. 32, May 23, 2008.

Palfreyman, D. (Ed.) (2001). *The Oxford Tutorial*. Oxford: Blackwell's

Appendix 1. Selected Findings from Case Studies

1. There was a faculty consensus that the assessment traits enhanced the pedagogy of the tutorials

Perhaps the most beneficial aspect of this study was having the tutorial students go through the rubric at the beginning of the year and discuss it with the instructors. The rubric led to wonderful discussions between instructors and students, where all parties were able to articulate some clear goals for the upcoming senior independent study project...I believe that this is an educationally sound practice—to give a student the rubric by which he/she will be assessed at the onset of the project, and to re-evaluate at mid-tutorial and at the end of the tutorial. If this process would take place with all tutorials, and honest feedback were to be given, then the grading process at the end of the year would not seem so mysterious (Pamela Pierce, Wooster)

My use of SAM-2 did influence my conduct of the tutorial. This revised format explicitly placed each student in a teacher role as either J. or C. was responsible for explaining a particular section of the book to the other student (Claudena Skran, Lawrence)

Working with SAM has absolutely realigned this perspective of mine. I am able to envision the goals of this tutorial (and others) both more broadly and more specifically. In other words, I understand the potential for student development to be well beyond the parameters I first envisioned, yet I have a much more sophisticated understanding of the varied levels of experience that make for that development (Stefanie Solum, Williams)

SAM-2 did have a few effects. First, it alerted me to the importance of setting up tutorials with three students of diverging disciplinary perspectives. I am now more likely to ask forward-looking questions at some point during the tutorial session. Inspired by the concept of “idea generation,” e.g., I am looking in tutorials for opportunities to ask students about the kinds of research they might develop to solve the problem they identified. I have increased my use of reflective writing in order to ask students to identify their emotional responses to their intellectual position—e.g., how they grapple with ambiguity and known biases (Patrick Schmidt, Macalester)

I found SAM-2 to be useful in assessing how the students progressed through the course of the tutorial and how well the tutorial succeeded as a whole. Carefully articulated learning outcomes can be a useful tool, but there needs to be a more regular way of keeping students focused on them. Filling out the rubrics after each tutorial session would be one way to do this. I feel confident that, at the very least, I will consult them (rubrics) about how to assess students in future courses, both tutorial and non-tutorial (Christopher Nugent, Williams)

The students’ peer interaction did help me was in my ability to work with the traits. Their interactions gave me numerous examples of the traits –reviewing the transcripts and papers, they just jumped out at me – and made it possible for me to work easily with them. These rubrics are flexible enough to be useful for tutorial evaluation in all

disciplines at Lawrence. I think they should be made available to our faculty as a “roadmap” to tutorial learning outcomes and goals (Jerald Podair, Lawrence)

The (assessment) criteria used proved to be useful in organizing the structure of the course and, at times, the discussions...Overall the method used serve to free me up from the need to be constantly evaluating (or “grading) the discussions and the work ...I found it to be a useful tool for the encouragement of (the seemingly contradictory ideas of) discipline *and* creativity (Rob Neilson, Lawrence)

I see in this rubric a model for the liberal arts that has been influential on my teaching. I have become much less satisfied with my more lecture-based courses and, wherever possible, have sought to put tutorials into those courses. (Patrick Schmidt, Macalester)

I think that participation in SAM-2 was a net positive for the students as it made them aware of the cognitive skills they should be developing in addition to understanding the content in the tutorial. I anticipate that some part of these rubrics will be used for tutorial evaluation and perhaps broader institutional assessment on the Lawrence campus. I think that I might also coordinate some of traits in these rubrics into my evaluations for student presentation, which are assigned in many biology courses. (Ron Peck, Lawrence)

2. The faculty encouraged the development of traits in students

I encouraged this trait (complexity/uncertainty) in every part of the tutorial—the conversation, written comments on papers, private meetings. I gave advice to her partner (Stefanie Solum, Williams)

I tried to help them develop this trait [complexity] through the selection of work that covered a similar topic, African development, but from different points of view (Claudena Skran, Lawrence)

Independence (taking teacher role) was among the most influential traits, if not the most important. I often invoke the aphorism, “to teach is to be twice taught”, with my students, and so they have been primed to think how leadership in the classroom both evidence and solidifies one’s learning.

Part of the influence of this trait is in setting the role of the tutor. Though I remain poised to enter a discussion, and do enter as needed, the trait reminds me to let there be silence. Sometimes it requires a little theatrical touch, such as looking at my notes or something as if to emphasize that I am not going to jump in on the conversation any time soon. In the vacuum, I am more likely to get students to initiate new directions. (Patrick Schmidt, Macalester)

RE: Self-assessment. One question that I do use as a point of entry is to ask whether the students had an initial conclusion based on the tutorial question but before they had read the materials. Although in some cases students simply didn’t know what they thought until they got into the readings, more often students do have an instinctive reaction. But students frequently report that they were substantially altered in their views, in whole or

in part; I enjoy hearing them articulate just how they had changed and why. (Patrick Schmidt, Macalester)

Z. honed his argumentation skills in our many conversations about literature, philosophy and gender and he became much more adept at drawing from analyzing personal experiences to explore contemporary gender dynamics. I pushed M. to take on the challenge of investigating his claim running wild in the popular media that a Super Bowl victory for the Saints would help the city recover from Katrina. I can't take much credit for "teaching" these skills on connectivity to Z: he came to the IS project with them finely honed. What I was able to do, however, was ask questions that might have furthered the connections he was seeing or lack of connections he wasn't seeing. M. also started IS with these abilities, although he wasn't as cognizant of them at the beginning, but became more so, increasing his self-awareness, which I always like to see (Nancy Grace, Wooster)

I encouraged this trait (inquiring mind) by constantly pushing students to dig deeper into the motivations behind the arguments we read. I emphasized that understanding the arguments was only a first step. They also need to think about why particular thinkers made particular arguments....One way I encouraged the development of this trait (self-assessment skills) in the tutorial was to have students identify at least two problematic aspects of every paper they wrote...I do not feel that either students improved substantially in their ability to address their biases and predispositions on their own. At the same time, when either I, or their partner alerted them to possible biases in their arguments, they were very receptive to the idea and seemed will to rethink things." My primary means of encouraging this trait (generating ideas) in the tutorial is through continuously pushing students to move beyond the secondary readings to come up with their own interpretations of primary materials." (Christopher Nugent, Williams)

I tried to encourage this trait (complexity) through my "counter-punch questions" which ask students to be a little more explicit about what you mean or to clarify what you mean. Not necessarily questions to give you ideas but to ask you to examine the ideas you have just articulated (Jerald, Podair, Lawrence)

I will try to place a bit more of the burden of critique on the students. I hope this promotes self-assessment as well as a more refined awareness of the necessary verbal critiquing skills they will need in graduate school and beyond. N. said: "I thought me and J asked each other really tough questions about one another's work." I encouraged "taking intellectual chances" by saying "I would rather see a spectacular failure than a mediocre success" (Rob Neilson, Lawrence)

(Being double majors) students brought multiple perspectives and connectivity into the course. (Rob Neilson, Lawrence)

Re: taking teacher role. Part of their evaluation for each class included a "Presentation" component where I encouraged them to describe complex topics at a level appropriate to their audience (i.e., "teach" them). I wanted the students to know that they were the

experts at their chosen topic and it was their responsibility to tell us about that topic. Re: creativity. I pushed the students to suggest their own topics for assignments and carefully avoided any efforts for them to get me to “suggest” a topic for them. When students did demonstrate initiative, I noted that in their evaluation and rewarded them with a higher grade. When I felt that students chose topics specifically to make the assignment “easier,” I also noted that in their evaluation (Ron Peck, Lawrence)

Each assignment was focused on a particular topic, which was inherently complex. The students then had to find a primary literature article and fully tackle the analysis of this article. Usually, students then had to analyze original data. In essence, grappling with complexity many times was a requirement for the class (Ron Peck, Lawrence)

Re: self-assessment I directed the students to assess their claims of the authors of the papers and also assess the author’s and their interpretations of the data. Re: creativity I required students to generate possible future directions based on the author’s results. Encouraged them to go far beyond simply restating or even modifying the author’s proposals (Ron Peck, Lawrence)

3. Students integrated the rubrics into their learning goals

I am aware that the SAM-2 influenced my students. I even overheard them (including those who were not being assessed, but were shown the rubric) talking about how they were trying to demonstrate some of the traits on the SAM-2. I think the language and ambition of “taking on teaching roles” registered with them in particular because my course is designed so that they must assign the readings at the end of the course. They usually opt to do this as pairs of students, but their selection was explicitly influenced by the desire to set up a competition of ideas. Indeed, they were more deliberate than past sections of the course in trying to generate problems for the class to think through, rather than pick readings that one-dimensionally fleshed out a problem from a perspective with which they agreed. (Patrick Schmidt, Macalester)

Several items on the rubrics stayed with them throughout the year. Occasionally, they would make a comment to me such as: “Did you notice that I was taking the teacher role during that discussion? Or “Was that a probing question?” (Pamela Pierce, Wooster)

Filling out the rubrics, especially the baseline one, seemed to make both C. and J. more aware of what they should be trying to learn and accomplish in the tutorial (Claudena Skran, Lawrence)

C. was perhaps the most self-reflective, concluding in his second assessment that “doing this” [taking chances] helps me better understand everything that we discuss. C. included the following in an observational note: ‘the more we go in depth with the topic, the more I’m able to absorb it all and raise new ideas’ (Claudena Skran, Lawrence)

The most significant evidence I have for “Independence” is that Z. told me first semester that he wanted to spend an undetermined number of weeks reading books and discussing

them with him. He explained in our final taped tutorial that the entire process gave him control that he didn't have in non-tutorial classes (Nancy Grace, Wooster)

The fact that both received honors on their project is evidence of their abilities to deal with uncertainty and complexity at a high level. I'm pleased too, that the SAM rubric indicates that they were aware of this skill as a strength each possessed (Nancy Grace, Wooster)

In the Exit Interview my student, A. responded as follows: "For me the learning outcomes helped me the most with preparing for the paper. I exerted the most effort in trying to improve my writing technique, which included formulating sound and valid arguments, following an orderly structure, creatively connecting disparate ideas together, and rendering complex arguments simple to understand. As for the tutorial session, I thought the outcomes helped me focus on asserting ideas and questions." "A. also found the self-assessment useful stating that, "To some degree the self-assessment helped me mark which skills were developing and which skills needed more work." "A.'s papers generally dealt with much more complex ideas and had correspondingly complex arguments {while] both L's papers and her discussions tended to stay away from complex issues" (Christopher Nugent, Williams)

There was substantial encouragement of the development of this trait (multiple perspectives) from peer interactions. A. frequently brought in ideas from his work in classes on philosophy, psychology and religion. L. would bring in ideas from her work in math and sciences. (Christopher Nugent, Williams)

RE: ability to take teacher roles. MA. "I like inviting audience participation through questions, but you should have a plan to lead your audience to the desired answer if nobody gets it. You can do this by restating the question, providing hints, or asking progressively simpler questions to get your audience started on the right path." PK. "You also have a natural technique for physically pointing out the things on the slide as you are describing it and drawing on the board when you need to make the point. I think that shows you are acting as a "teacher" to explain something to the audience." LY. "I could tell you made the effort to slow down and clearly describe complex concepts...you also oriented the audience to your slides and almost always provided a clear 'take-home' message for each slide" (Ron Peck, Lawrence)

Some nuance was given to this assessment of the tutorials, in a further discussion among the students, and it is worth reporting here, following a question I asked about the most important dimensions of the SAM-2:

E.: I think learning to argue is a big one.

P.M.: And the self-assessment, and self-questioning. I was just thinking those three, and then in a row seemed to really embody what it meant for me, it was a real experience...

E.: I feel like with every time I would read this tutorial prompt, I would have a knee-jerk stance that I was going to take, and then as I went through the readings, I would be like, well, why do I think that? And sometimes...usually I would, usually I would make with

the same thesis that was maybe doctored a little from my initial reaction to the prompt, but sometimes it changed some. (Patrick Schmidt, Macalester)

4. The rubrics discriminated between students who varied in performance

At the onset of the project, it appeared that the rubric was motivating the weaker of my two students...it would seem natural to ask whether the rubric provides additional motivation to a more 'average' student. After a few initial meetings, R. was able to lead the discussion and initiate topics she wanted to discuss. K. on the other hand, needs a lot of guidance, and we had to prod her to keep moving (Pamela Pierce, Wooster)

I noticed a real difference in the way these two students, M. and L., discussed their goals and responded to my questions. M. was thoughtful and slower to formulate answers about what her learning goals were (Stefanie Solum, Williams)

J. started out asking fairly high-level questions. C., however, struggled with high-level questions and initially asks rather basic questions (Claudena Skran, Lawrence)

Z. began as much more independent than M in terms of setting topics, asking questions to originate new topics, summarizing discussion and assessing ideas. Both M. and Z. functioned at a high level in terms of self-assessment, although M. was more vocal about this than Z (Nancy Grace, Wooster)

Under SAM-2 I found myself particularly aware that differences in individual students seemed to have more effect on their improvement over the course of the semester than did awareness of the learning outcomes... L. seemed to come to our readings with more biases and predispositions toward the material and subject matter than most students. (Christopher Nugent, Williams)

Both students dealt with complexity and uncertainty with great acumen. J., however, exhibited astuteness in her undertaking by examining a long dead revolutionary Surrealist poet's writings and drawings to create an extraordinary sculpture installation (Rob Neilson, Lawrence)

RE: taking teacher role. All students showed improvement. MA. Immediately took my feedback and then refined this aspect over the course of the term. PK and LY improved, but to smaller degrees and more sporadically during the term. RE: Self-assessment. Best students seemed to already have the ability and awareness to properly evaluate evidence, but I struggled to see or coax development of this trait in the weakest student (PK) (Ron Peck, Lawrence)

RE: Connectivity. This was the one dimension where L. reported a two-step gain (from 3 to 5) over the semester. She wrote effusively that "I have definitely improved my ability to synthesize bits of information and theories/established ideas. The sort of 'detective work' we have to do with the readings we are given, and finding connections in order to synthesize an argument, is really helpful." M. reported a gain in this dimension (from 3

to 4), describing connectivity as “inherent in tutorials”. S. was already strong in this dimension, and it was the one area where he allowed himself a score of “5” at the start of the semester. The background in interdisciplinary work, he wrote, “has strongly encouraged me to cultivate” this trait—which he likened to “intellectual promiscuity”. So, yes, I do think they improved in this dimension, and think it is both the product of tutorials and the particular structure of this tutorial course, in that it asks for both disciplined work and interdisciplinary interaction. (Patrick Schmidt, Macalester)

5. The rubrics were not useful in non-tutorial courses, but could be applied in small seminars

The rubrics will have less-utility in non-tutorial courses, except as ideas to keep in mind as basic pedagogical goals. I can see the rubrics helping change, however, perhaps to develop alternatives to the graduate-school seminar model many of us continue to use for undergraduates (Stefanie Solum, Williams)

The larger one (45 students) was primarily a lecture class, so it would have been difficult for me to gauge, outside of exams and papers, the progress of students on the traits covered by the rubrics (Claudena Skran, Lawrence)

I believe the SAM-2 Method is far less effective a tool for evaluating students in non-tutorial courses. Perhaps it is the size of any given section, but inevitably I found myself searching (and at times reaching) for examples that illustrated one of the prescribed traits...the lack of any sustained conversation on the part of any one student during a typical class session made it difficult to ascertain the amount or depth of growth in any one particular trait well enough for a truly meaningful quantification (Rob Neilson, Lawrence)

My non-tutorial courses are not designed to push and to reveal the SAM-2 dimensions. I engage with students during (large) class, often Socratically, but my questioning is far more purposive at the time and I do not get the chance to listen as carefully as I do in tutorial (when other students are just as likely to be doing the questioning of their peers). Tutorials provide more space in which students can reveal their ambiguities. They can take more risks. I can observe changes in position over the course of the semester. (Patrick Schmidt, Macalester)

I feel extensive use of the rubrics might work best in something like a senior or junior seminar in which students all come in with some familiarity with the subject matter and have a base from which to work (Christopher Nugent, Williams)

Appendix 2. Development of Learning Outcome Rubrics: SAM-1

Introduction

A Shared Assessment Method (SAM-1) Workshop was held in fall 2008 for the purpose of discussing tutorial frameworks among participating faculty members. Six of the eight faculty participants were in attendance, including the two co-principal investigators and a guest contributor from St. John's College – Annapolis. Faculty not in attendance received a videotape of the workshop to enhance understanding of the rubric generation process.

The workshop consisted of a two-day discussion regarding desired learning outcomes of tutorials and the measurable traits that are encompassed by those outcomes. The ability to observe and the faculty support necessary for students to develop the desired traits were at the forefront of discussion.

Learning Outcomes

Independent Thinker

The workshop included a presentation by Rob Beck, co-principal investigator, outlining notable literature on the assessment of student outcomes in liberal education and on desirable traits/dispositions of students in tutorial settings. Specifically, a study of the Oxford tutorial found that independence/thinking for oneself in learning was a central outcome. Beck (2007) laid the groundwork for Independent Thinker to be considered as a desired student outcome to be assessed.

Eight traits and thirty-five sub-traits related to the Independent Thinker learning outcome were discussed by faculty for the purpose of sharing and determining which sub-traits had been observed in previous tutorials or whether observation was possible of each and how faculty should support the development of the traits or activities in a tutorial setting. Traits that were deemed unobservable were not researchable, while those that were observable were considered by the group and discussed. The co-principal investigators considered all feedback from faculty at the workshop and ultimately pared down the Independent Thinker learning outcome to include six observable traits (two with two sub-traits each and one with three sub-traits) that would be evaluated in tutorial sessions for the purposes of this study:

Traits	Description
Independence/dependence - how much support the student needs	a) dependence - the need for frequent oral and written instructions that are explicitly provided to carry out tutorial activities, e.g., ways to do research, writing, etc.; b) independence - the ability to take teacher roles during discussions by setting topics, asking questions to originate new topics, summarizing discussion and assessing ideas
Developing an inquiring mind - questioning ideas and self-questioning	a) asks high-level questions, such as why-questions; b) meticulous in articulating questions and stating them, as far as possible, clearly and precisely; c) acquires the habit of puzzling over ideas, that is, asks oneself questions
The Interdependent Thinker	takes the perspective of others, considers alternative ideas during discussions, and demonstrates mental flexibility by modifying one's ideas and changing one's mind
Acquiring self-assessment and assessment skills	a) self-assesses one's claims and arguments; b) uses assessment/feedback to improve one's work and to seek assessment/feedback on one's work
Skeptical orientation	adopts a posture of doubt and uncertainty about claims and arguments including one's work and the work of peers
Learning to argue	argues effectively by making conceptual claims backed with supporting theory and evidence

Intellectual Maturity

A great deal of discussion by faculty at the workshop concerning tolerance of ambiguity, intellectual humility, risk taking, recognizing mistakes, and self awareness was determined to fall under the broad learning outcome of Intellectual Maturity. Ways in which these traits can be observed and supported in tutorials were discussed. Four measurable traits (one with 2 sub-traits) were chosen by the co-principal investigators to represent this learning outcome:

Traits	Description
Self-awareness of intellectual biases and prejudices	the ability to recognize biases in thinking and predispositions to make certain judgments, shares or communicates critique of one's biases and prejudices
Tolerance of ambiguity	accepts and welcomes "gray areas" and complexity, not taking refuge in certitudes and being able to embrace uncertainty
Intellectual humility	recognition and understanding that I don't know all the answers, ideas are subject to revision and change
Risk-taking	a) the ability to state positions and arguments without worry of saying something wrong, willingness to make mistakes, risk failure in discussions, writings, and/or presentations; b) willingness to reveal that thinking has changed

Creativity

Creativity was presented as a desired outcome of tutorials stemming from The Five Colleges of Ohio Creative and Critical Thinking Project which developed tools to assess creativity and critical thinking outcomes in a liberal arts setting. One of the faculty participants involved in the Five Colleges of Ohio project was also a participant in this Teagle Tutorial Assessment study.

Thirteen traits were considered by the faculty at the workshop as possible evaluation criteria by discussing the observable nature of each and ways in which faculty should support the

development of the traits. The co-principal investigators considered feedback from the faculty participants and ultimately chose five traits to be evaluated in tutorial sessions concerning the learning outcome of Creativity:

Traits	Description
Idea generation	generates new ideas, variations of or alternatives to solving problems, a novel way of analyzing or re-conceptualizing a topic or idea in the context of what the student knows and understands, interesting and creative restatements of others' ideas, unusual ideas, interesting theories
Curiosity	the desire to learn or know more, ability to become absorbed in the topic, discovers a new line of inquiry or question of a topic and wishes to persist and sustain in exploring the topic
Multiple perspectives in problem solving	sees a problem from multiple perspectives, compares and contrasts approaches, uses multiple disciplines
Connectivity	ability to bring together disparate bits of information, makes connections between already established ideas or theories, connects disciplines
Divergent thinking	ability to go against the grain of the usual or expected, thinking outside the box, originality, ability to extend an idea

Tutorial Course Organization and Evaluation: SAM-1

Each tutorial course enrolled one or more students to engage in discussions, write papers, and if required, present their work. A total of 20 tutorial students participated in SAM-1.

Learning outcome rubrics were used for the assessment of discussions, papers, and/or presentations at three times during the tutorial. These three assessment periods were referred to as the baseline, midpoint, and final assessments. Students were also instructed to complete the rubrics to self-assess themselves during assessment periods. Each assessment period was audio recorded and transcripts were created.

The following procedural timeline indicates the evaluation points during each tutorial course:

- Orientation interview – to describe the goals and methods of the course, review the informed consent form including compensation for the student. Discussion and use of the learning outcome rubrics to be used in assessing student performance. This interview is not recorded.
- Baseline assessment – occurs during the second or third week of class and serves as the baseline measure of the learning outcomes. The rubrics are used in assessing the discussion that occurs in one class and any written work that occurs during this window of time. This discussion is recorded and transcribed.

- Midpoint assessment – occurs during the fifth (for terms) to eighth (for semesters) week of class and serves as the mid-point measure of outcomes. The same procedures are followed as for the baseline assessment.
- Final assessment – occurs during the last couple weeks of the tutorial and serves as the end-point measure of the learning outcomes. The same procedures are followed as for the baseline assessment.
- Exit interview – to discuss progress and development during the tutorial. This interview is recorded and transcribed.

Rubrics

All faculty participants reached consensus concerning the three learning outcomes for SAM-1, Independent Thinker, Intellectual Maturity, and Creativity as well as the traits of those outcomes. The co-principal investigators developed rubrics by which the faculty would be able to measure the traits in both discussion with and written work by the student. Each trait was to be evaluated on a 5-point scale (1=Never to 5=Very Frequently) concerning how often a student exhibited that trait. Faculty were also given the opportunity to record detailed qualitative observations they made of the traits exhibited in their tutorial student(s).

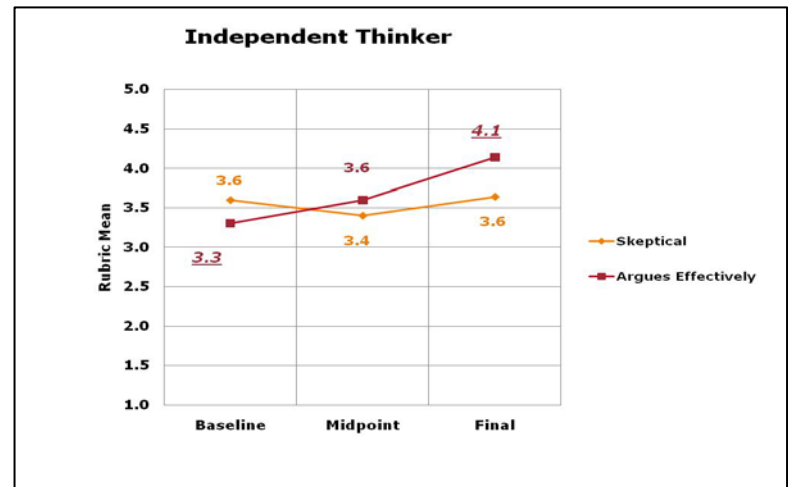
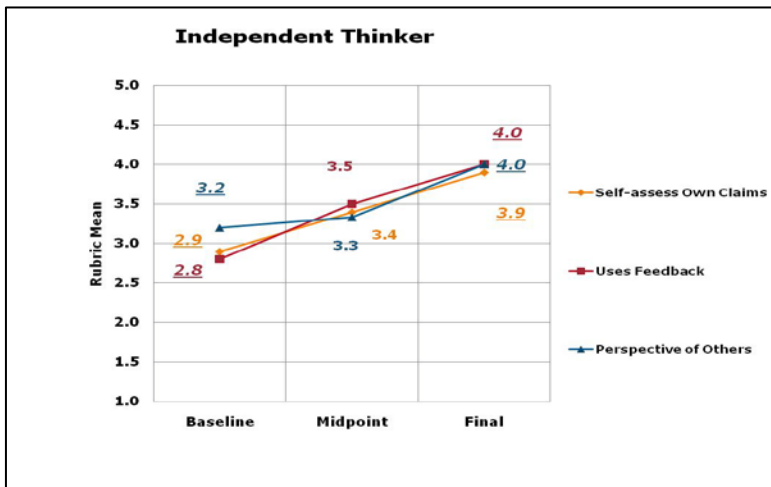
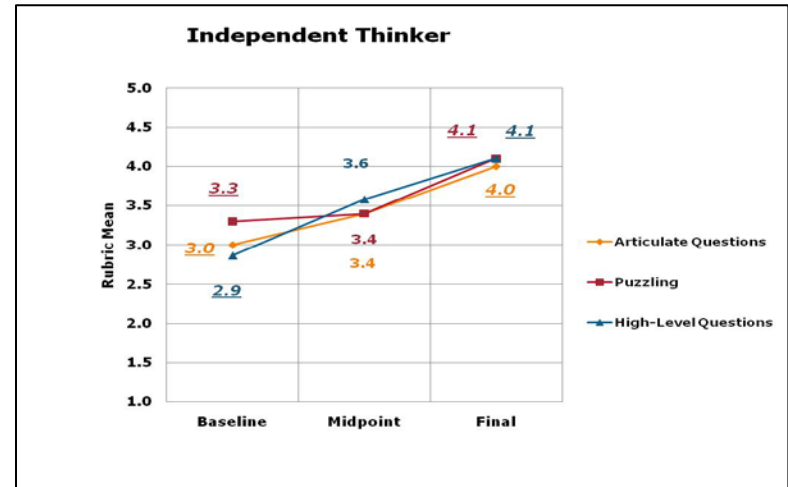
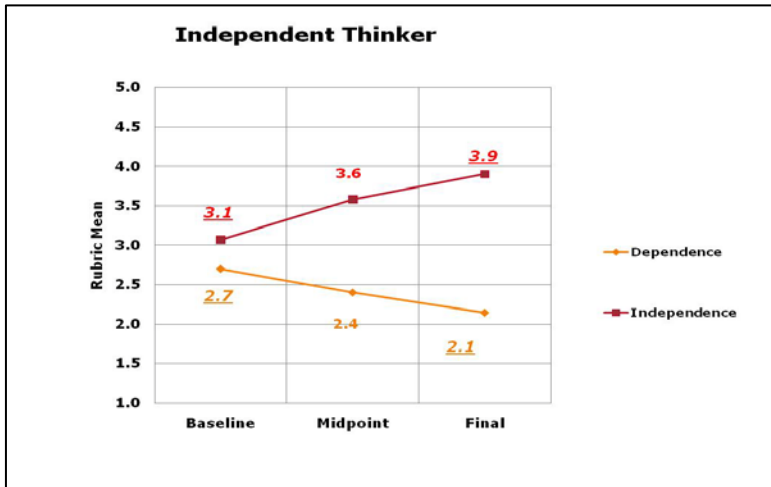
SAM-1 Rubric Analysis

Appendix 3 and 4 contain the results of the faculty rubric analysis for the traits of three tutorial learning outcomes for SAM-1: Independent Thinker, Intellectual Maturity, and Creativity. The graphs show the mean rating for each trait across three time points – baseline, midpoint, and final. All of the traits associated with Independent Thinker (Appendix 3) except *skeptical* showed statistically significant improvement from baseline to final as indicated by the underlined means. It is interesting to note that as *dependence* significantly decreased from baseline to final, *independence* significantly increased. Also, for the other traits (except *skeptical*), there is not a large range in the means at each time point. The baseline means are between 2.8 and 3.3, the midpoint means are between 3.3 and 3.6, and the final means are between 3.9 and 4.1. In Appendix 4 it can be seen that the mean rating for each trait of Intellectual Maturity and Creativity also significantly improved from baseline to final. Also of interest is the baseline mean of 3.6 for *curiosity* is the highest starting point of all other traits

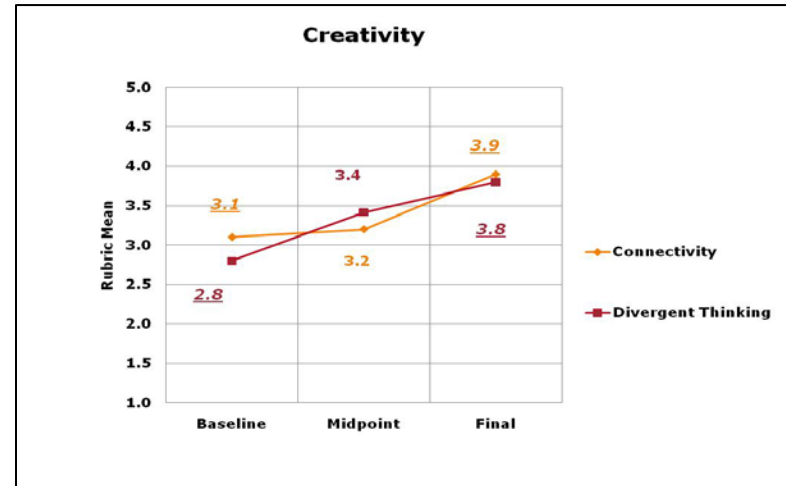
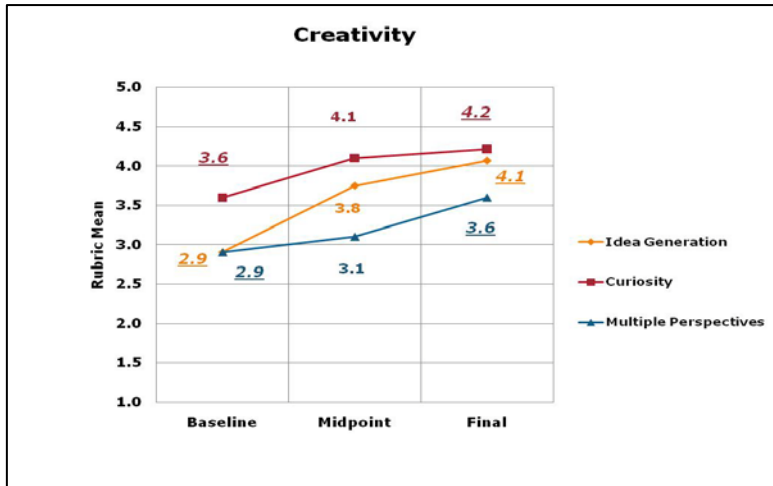
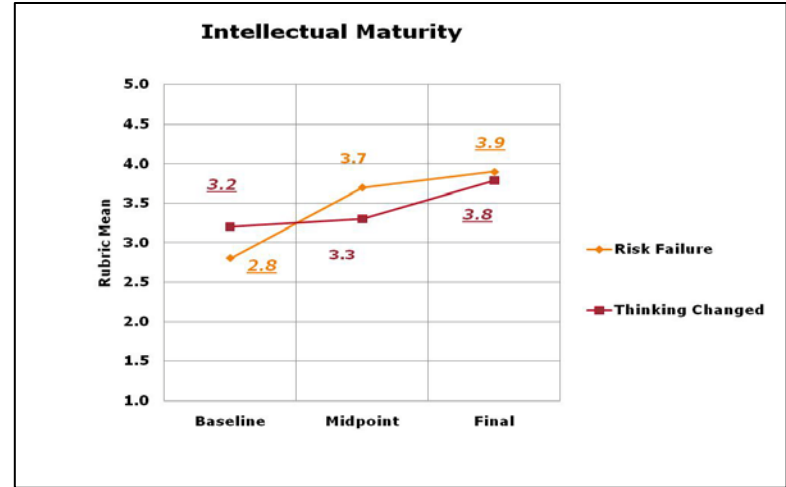
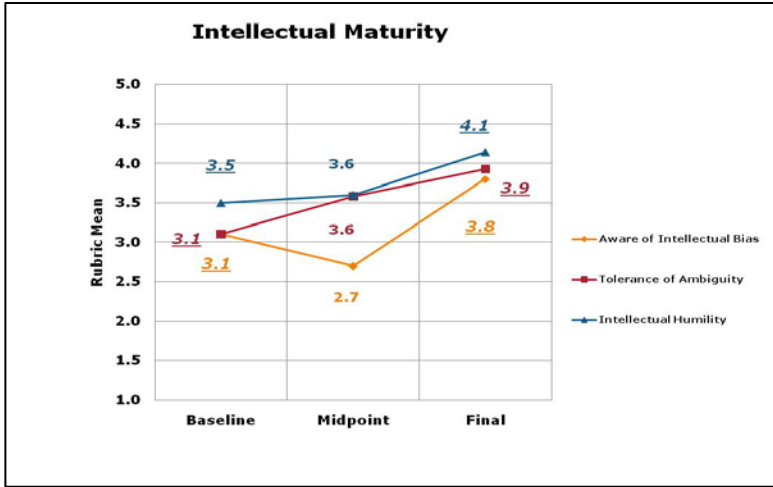
except *skeptical*, which has the same baseline mean of 3.6. This suggests that students enter the tutorial with a relatively high desire to learn or know more and become absorbed in a topic (*curiosity*) and high level of doubt and uncertainty about claims and arguments (*skeptical*). Finally, the traits that show the largest improvement from baseline (2.9) to final (4.1) are *high-level questions* and *idea generation* suggesting the tutorial is an effective setting to develop these traits.

Appendix 5 compares the faculty and student means for the learning outcome traits. This is a limited comparison because only eight students completed the baseline, midpoint, and final rubrics and the corresponding faculty ratings of those students who did not complete the rubrics are not included in the calculation of the faculty means in this table. Based on the students' self-ratings, all the traits showed improvement from baseline to final. Statistically significant improvement (indicated by superscript "a") in the traits from baseline to final occur for *dependence*, *independence*, *articulate questions*, *self-assess own claims*, *argues effectively*, and *aware of intellectual bias*. There is remarkable similarity between faculty and students on the mean ratings for most of the traits. Statistically significant differences (indicated by superscript "b") in the mean ratings between faculty and student only occur at baseline for *uses feedback* and *multiple perspectives* and at final for *risk failure*.

Appendix 3. Faculty Rubrics for SAM-1: Independent Thinker



Appendix 4. Faculty Rubrics for SAM-1: Intellectual Maturity and Creativity



Appendix 5. Faculty and Student Baseline, Midpoint, and Final Rubric Means for SAM-1

Trait	Faculty Means (N=8)			Student Means (N=8)		
	Baseline	Midpoint	Final	Baseline	Midpoint	Final
INDEPENDENT THINKER						
Dependence	2.8	2.7	2.3	2.6 ^a	2.4	2.1 ^a
Independence	3.1	3.4	4.1	2.9 ^a	3.0	3.9 ^a
Articulate Questions	2.9	3.3	4.0	2.9 ^a	3.2	3.6 ^a
Puzzling	3.6	3.7	4.3	3.9	3.4	4.1
High-Level Questions	3.0	3.9	4.1	3.3	3.2	3.9
Self-Assess Own Claims	3.1	3.7	4.4	3.4 ^a	4.2	4.5 ^a
Uses Feedback	3.0 ^b	3.9	4.3	3.9 ^b	4.6	4.3
Skeptical	4.0	3.4	3.5	3.8	3.6	3.9
Perspectives of Others	3.5	3.3	3.8	3.6	3.8	4.3
Argues Effectively	3.3	3.9	4.3	3.1 ^a	3.8	4.0 ^a
INTELLECTUAL MATURITY						
Aware of Intellectual Bias	3.3	3.2	3.8	2.9 ^a	4.0	4.0 ^a
Tolerance of Ambiguity	3.0	3.7	3.8	3.5	4.2	3.9
Intellectual Humility	3.6	3.7	4.1	3.9	3.8	4.1
Risk Failure	3.1	3.9	4.4 ^b	3.0	3.0	3.5 ^b
Thinking Changed	3.3	3.3	3.5	3.4	3.8	3.9
CREATIVITY						
Idea Generation	3.0	4.0	4.1	3.6	3.4	4.1
Curiosity	3.6	4.3	4.3	3.8	3.4	4.1
Multiple Perspectives	2.7 ^b	3.3	3.5	3.6 ^b	3.8	3.8
Connectivity	3.1	3.6	4.3	3.4	3.8	3.9
Divergent Thinking	2.8	3.9	3.9	3.0	2.8	3.8

^aBaseline and final student means significantly different at $p < .05$

^bBaseline or final faculty and student means significantly different at $p < .05$